# A Markov Chain Monte Carlo procedure to generate Revealed preference consistent datasets

Thomas Demuynck\*

August 4, 2020

#### Abstract

This paper presents a Markov-Chain-Monte-Carlo (MCMC) procedure to sample uniformly from the collection of datasets that satisfy some revealed preference test. The MCMC combines a Gibbs-sampler with a simple hit and run step. It is shown that the MCMC has the uniform distribution as its unique invariant distribution and that it converges to this distribution at an exponential rate.

**Keywords:** Revealed preference, Markov Chain Monte Carlo, Bronars Power **JEL codes:** C63, D11

#### 1 Motivation

Revealed preference theory is a versatile nonparametric method that allows to test for rational decision behaviour given some finite set of choices. It's main attractiveness comes from the fact that it is entirely nonparametric, i.e. it does not rely on auxiliary parametric assumption imposed on the utility function. This feature makes that revealed preference methods are frequently used for testing the hypothesis of utility maximizing behaviour. In his seminal contribution, Afriat (1967) showed that a finite data set on observed prices and consumed bundles is consistent with utility maximizing behaviour if and only if it satisfies

<sup>\*</sup>ECARES, Université Libre de Bruxelles. Avenue F. D. Roosevelt 50, CP 114, B-1050 Brussels, Belgium. E-mail: thomas.demuynck@ulb.be. Thomas Demuynck acknowledges financial support by the Fonds de la Recherche Scientifique-FNRS under grant nr F.4516.18

the Generalized Axiom of Revealed Preference (GARP).<sup>1</sup> Nowadays there exists a wide variety of revealed preference tests for various models of decision making.<sup>2</sup>

This paper adds to the revealed preference toolbox by providing a Markov Chain Monte Carlo (MCMC) procedure to sample uniformly from the collection of datasets that satisfy some revealed preference test. In the main text, we will focus on obtaining random data sets that satisfy GARP, but we also discuss other revealed preference tests, like the Weak Axiom of Revealed Preference (WARP) (Samuelson, 1938), the Homothetic Axiom of Revealed Preference (HARP) (Varian, 1983) and the Quasi-Linear Axiom of Revealed Preference (QLRP) (Brown and Calsamiglia, 2007). In addition, we show that the method can easily be modified to generate datasets that are nearly consistent with a particular revealed preference test.

The possibility to sample from the set of all datasets consistent with a revealed preference test can be useful for a variety of reasons. First of all, it may help to conduct a so called Bronars power analysis (Bronars, 1987). The Bronars power gives the probability that a randomly generated data set violates the revealed preference test. It is usually calculated using a Monte Carlo procedure by generating a large number of random datasets and computing the fraction of these datasets that violate the revealed preference test. As usual with Monte Carlo integration, the (relative) accuracy of this procedure depends on the number of randomly generated datasets and on the size of the target. If the revealed preference test is very strict, then only a very small number of the randomly generated datasets will satisfy the revealed preference test. In such cases, the obtained estimate of this target will be quite unreliable (in relative terms), unless one generates a very large number of random data sets. Doing this becomes computationally very demanding. In section 5 we will show how our MCMC method can help in such instances.

Second, the MCMC procedure may be useful to compare the strictness of different revealed preference tests, especially if they are nested, i.e. if consistency with one test automatically implies consistency with the other test. Instead of computing the power separately between the two tests, we may think of first generating random data sets that satisfy the weaker test and only check consistency with the more stringent test on these datasets (see Heufer (2013) for a similar argument). Section 5 also illustrates this feature.

Finally, the MCMC procedure might be useful in the development of (non-parametric)

<sup>&</sup>lt;sup>1</sup>See also Varian (1982) who actually coined the term GARP. Afriat (1967) extended earlier work of Samuelson (1938) and Houthakker (1950).

<sup>&</sup>lt;sup>2</sup>See Chambers and Echenique (2016), Crawford and De Rock (2014) and Demuynck and Hjertstrand (2020) for recent overviews of the literature.

statistical tests of rational or irrational consumer behaviour (e.g. to test for negativity or symmetry of the Slutsky matrix). Consider a setting where one has an (asymptotic) test to verify or reject the rationality of consumer behaviour. In order to evaluate the correctness of this test, or in order to analyse its performance in finite samples, it may be useful to see how the test performs using a Monte Carlo exercise. To do this, it might be necessary to simulate choice behaviour from a rational consumer. Our MCMC method allows to efficiently generate data that fit such description and could therefore be used to in the development of such statistical tests.

Our MCMC algorithm is based on a hit-and-run (H&R) procedure, which is a well known method for generating samples from the uniform distribution on a bounded region in Euclidean space (Bélisle, Romeijn, and Smith, 1993). Given a starting vector x in the region of interest, H&R first picks a random direction,  $\delta$ , uniformly from the unit hypersphere. Next it determines the possible values of  $\lambda$  such that  $x + \lambda \delta$  is still in the target region. Finally, it picks a value  $\lambda^*$  uniformly from this range and set the next value of the chain equal to  $x + \lambda^* \delta$ .

The H&R algorithm is easy to implement if the target region is a convex set, as then the range of possible  $\lambda$ -values is an interval whose bounds can be computed fast using a binary search procedure. If the target region is not convex, then determining this set is much harder. Unfortunately, the collection of GARP consistent datasets is not always a convex set, so implementing the default H&R algorithm is not straightforward. However, as we will show in the paper, if we keep all observations but one fixed, then the set of all GARP consistent datasets does become convex. This leads to the following 'Gibbs'-adjustment of the standard H&R method: (i) draw an observation at random and (ii) perform H&R on this observation only, keeping the quantities for all other observations fixed. As shown in this paper, this modified H&R algorithm gives a Markov Chain that has the uniform distribution as its unique invariant distribution. Also, this chain is uniformly geometric ergodic, which means that it approximates the uniform distribution at an exponential rate.

The paper closest to the current one is by Heufer (2013). In his paper, Heufer (2013) develops a Monte Carlo procedure that generates random GARP consistent datasets. His procedure constructs the random data set observation by observation. For each additional observation, his algorithm guarantees that adding the new observation does not lead to a GARP violation. In particular, a new observation is chosen by applying a H&R step to the region of possible values for the new observation that are GARP consistent with the previously retained observations. Although Heufer (2013)'s algorithm guarantees to

generate a random GARP-consistent dataset, the underlying distribution of the constructed datasets is not easily analysed. The algorithm in this paper has the advantage of belonging to the widely studied MCMC class of algorithms. This feature allows the usage of known methods to demonstrate desirable convergence properties.<sup>3</sup> Further, we show that the MCMC algorithm is easily extendible to other revealed preference tests.

Section 2 gives the necessary notation and definitions. Section 3 describes the main MCMC algorithm. In Section 4, we show how this algorithm can easily be modified to generate datasets consistent with other revealed preference tests. Section 4 contains an illustration. All proofs and technical details are in the Appendix.

#### 2 Setting

We consider a setting with k + 1 goods  $\{0, 1, \ldots, k\}$ . Throughout this paper, we consider a fixed set of price vectors  $\mathbf{p} = (p_t)_{t \leq T}$  where each  $p_t$  is a vector of k + 1 strictly positive prices  $p_t = (p_{t,0}, p_{t,1}, \ldots, p_{t,k}) \in \mathbb{R}^{(k+1)}_{++}$  corresponding to the k + 1 goods. For convenience, we normalize prices such that the total budget equals one and we assume that all price vectors are distinct.<sup>4</sup> A dataset  $\mathbf{q} = (q_t)_{t \leq T}$  consists of T bundles  $\mathbf{q} = (q_t)_{t \leq T}$  where each bundle consists of k + 1 quantities  $q_t = (q_{t,0}, q_{t,1}, \ldots, q_{t,k}) \in \mathbb{R}^{k+1}_+$  for the various goods. As prices are normalized, we have the restriction that total expenditures add up to one:

$$1 = p_t \cdot q_t = \sum_{i=0}^k p_{t,i} q_{t,i}.$$

We call T the size of the data set. The underlying idea is that  $q_t$  is the consumed bundle at observation t.

The observation t is defined to be *directly revealed preferred* to the observation v if  $q_v$  could have been bought at observation t:

$$p_t \cdot q_t = 1 \ge p_t \cdot q_v$$

We also write this as t R v and call R the revealed preference relation. The observation t is

<sup>&</sup>lt;sup>3</sup>See for example the key references of Nummelin (1984) and Meyn and Tweedie (1993).

<sup>&</sup>lt;sup>4</sup>If the total budget at observation t, say  $m_t$ , would differ from one, we can normalize the prices by dividing every price  $p_{t,i}$  by the income  $m_t$ .

strictly directly revealed preferred to the observation v if:

$$1 > p_t \cdot q_v$$

If so, we write t P v. We call P the strict revealed preference relation.

A dataset  $\mathbf{q} = (q_t)_{t \leq T}$  is said to satisfy the *Generalized Axiom of Revealed Preference* or GARP if the revealed preference relations R, P do not contain a cycle. In particular: for all sequences of distinct observations  $t_1, \ldots, t_M$ :

 $t_1 R t_2 \ldots R t_M$  implies not  $t_M P t_1$ .

Afriat (1967) showed that a dataset  $\mathbf{q} = (q_t)_{t \leq T}$  with prices  $\mathbf{p} = (p_t)_{t \leq T}$  is rationalisable by a locally-nonsatiated utility function  $u : \mathbb{R}^{k+1}_+ \to \mathbb{R}$  if and only if GARP is satisfied. Here rationalisability means that for all  $t \leq T$  and all  $q \in \mathbb{R}^{k+1}_+$ :

if 
$$1 \ge p_t \cdot q$$
 then  $u(q_t) \ge u(q)$ .

In words, if the bundle q was avalable (but not chosen) at observation t, then the utility at the chosen bundle,  $q_t$ , must be at least as high as the utility at q.

For the remainder of the paper, it will also be convenient to express datasets in terms of shares instead of quantities. For each consumption bundle  $q_t = (q_{t,0}, q_{t,1}, \ldots, q_{t,k})$  that satisfies the budget condition,  $p_t \cdot q_t = 1$ , we correspond a share vector  $s_t = (s_{t,0}, s_{t,1}, \ldots, s_{t,k})$  such that:

$$s_{t,i} = p_{t,i} \, q_{t,i}.$$

The number  $s_{t,i}$  give the fraction of the total budget spent on good i (remember that we normalized budgets to one). Shares vectors  $s_t = (s_{t,0}, \ldots s_{t,k})$  belong to the k dimensional unit simplex  $\Delta^k$ .

$$\Delta^{k} = \left\{ s \in \mathbb{R}^{(k+1)}_{+} : \sum_{i=0}^{k} s_{i} = 1 \right\}.$$

Depending on the setting, we will interchange frequently between the share and bundle representation. As such, a dataset will be denoted by  $\mathbf{s} = (s_t)_{t \leq T}$  or  $\mathbf{q} = (q_t)_{t \leq T}$  interchangeably. In this sense, we will say that the dataset  $\mathbf{s} = (s_t)_{t \leq T}$  satisfies GARP if the corresponding collection of consumption bundles  $\mathbf{q} = (q_t)_{t \leq T}$  satisfies GARP.

A dataset  $\mathbf{s} = (s_t)_{t \leq T}$  consists of T vectors in  $\Delta^k$ , so it is an element of  $(\Delta^k)^T$ . This means that  $(\Delta^k)^T$  coincides with the collection of all possible datasets. The collection of

all datasets **s** that satisfy GARP will be denoted by  $\Omega_{GARP}$ .

$$\Omega_{\text{GARP}} = \left\{ \mathbf{s} \in (\Delta^k)^T : \mathbf{s} \text{ satsifies GARP} \right\}.$$

Our aim is to sample datasets uniformly from the set  $\Omega_{\text{GARP}}$ . Towards this end, define by  $\mu$  the uniform measure on  $(\Delta^k)^T$ , i.e. for a measurable set A,  $\mu(A)$  gives the size of A relative to  $(\Delta^k)^T$ .<sup>5</sup> Next, define the uniform probability measure  $\nu$  on  $\Omega_{\text{GARP}}$ : for all measurable sets  $A \subseteq (\Delta^k)^T$ .<sup>6</sup>

$$\nu(A) = \frac{\mu(A \cap \Omega_{\text{GARP}})}{\mu(\Omega_{\text{GARP}})}$$

The next section will present a Markov Chain Monte Carlo method to generate samples from  $\nu$ .

#### 3 The MCMC algorithm

The MCMC procedure combines a Gibbs sampler with a H&R step. At every iteration, one observation  $t \in \{1, \ldots, T\}$  is picked at random and the share vector  $s_t$  for observation t is updated. The update is performed using a H&R step over the set of all share vectors that preserve GARP-consistency.

Consider a dataset  $\mathbf{s} = (s_v)_{v \leq T}$  and a share vector  $\tilde{s}_t \in \Delta^k$ . We use the following notation to present the new dataset that is obtained by replacing  $s_t$  by  $\tilde{s}_t$ :

$$(\tilde{s}_t, s_{-t}) = (s_1, \dots, s_{t-1}, \tilde{s}_t, s_{t+1}, \dots, s_T),$$

For  $\mathbf{s} = (s_v)_{v \leq T} \in \Omega_{\text{GARP}}$ , we define:

$$\mathcal{P}(s_{-t}) = \left\{ \tilde{s}_t \in \Delta^k : (\tilde{s}_t, s_{-t}) \in \Omega_{\text{GARP}} \right\}.$$

The set  $\mathcal{P}(s_{-t})$  contains all share vectors  $\tilde{s}_t$  such that the data set  $(\tilde{s}_t, s_{-t})$  satisfies GARP.

The following Lemma shows that this set is convex.

**Lemma 1.** Let  $\mathbf{s} = (s_v)_{v \leq T} \in \Omega_{GARP}$ , then for all  $t \leq T$ ,  $\mathcal{P}(s_{-t})$  is a non-empty convex set.

<sup>&</sup>lt;sup>5</sup>See Appendix A on how  $\mu$  is constructed in a sound measure-theoretic sense.

<sup>&</sup>lt;sup>6</sup>Lemma 6 in Appendix C demonstrates that  $\mu(\Omega_{\text{GARP}}) > 0$ , so this is well defined. Also, as  $\Omega_{\text{GARP}}$  is defined by a finite number of weak and strict inequalities, it is  $\mu$ -measurable.

Let us now give a description of the H&R step. Fist, we need to get a random direction. Let  $\partial D$  be the set of all directions on  $\Delta^k$ :

$$\partial D = \left\{ \delta \in \mathbb{R}^{k+1} : \|\delta\| = 1 \text{ and } \sum_{i=0}^k \delta_i = 0 \right\}.$$

The set  $\partial D$  contains all vectors whose elements sum to zero and are on the surface of the (k+1) dimensional unit sphere. Observe that if  $s \in \Delta^k$  and  $\delta \in \partial D$ , then for all  $\lambda \in \mathbb{R}$ :  $\sum_{i=0}^k (s_i + \lambda \delta_i) = 1$ . Algorithm 1 shows how to draw an element  $\delta$  uniformly from  $\partial D$ .<sup>7</sup>

<b>Algorithm 1</b> Sample $\delta$ uniformly from $\partial D$			
Require: k			
1: Draw $k + 1$ i.i.d. standard normally distributed variables $x = (x_0, \ldots, x_k)$ .			
2: Compute $y = (y_0, \ldots y_k)$ where $y_i \leftarrow x_i - \sum_{j=1}^K \frac{x_j}{K}$			
3: Compute $\delta = (\delta_0, \dots, \delta_k)$ where $\delta_i \leftarrow \frac{y_i}{\ y\ }$ .			
4: return $\delta$			

For a given share vector  $s \in \Delta^k$  and direction  $\delta \in \partial D$ , let  $\Lambda(s, \delta)$  contain all numbers  $\lambda$  such that  $s + \lambda \delta$  is in  $\Delta^k$ .

$$\Lambda(s,\delta) = \{\lambda \in \mathbb{R} : s + \lambda \delta \in \Delta^k\}.$$

This set can easily be found. In particular, for all goods i = 0, ..., k, we have the bounds:

$$0 \le s_i + \lambda \delta_i \le 1$$

If  $\delta_i > 0$ , this gives the bounds:

$$\frac{-s_i}{\delta_i} \le \lambda \le \frac{1-s_i}{\delta_i}$$

If  $\delta < 0$ , we obtain the bounds:

$$\frac{1-s_i}{\delta_i} \le \lambda \le \frac{-s_i}{\delta}.$$

This gives:

$$\Lambda(s,\delta) = [\underline{\lambda},\lambda],$$

<sup>&</sup>lt;sup>7</sup>See also Muller (1959).

where:

and

$$\underline{\lambda} = \max\left\{\max_{i}\left\{\frac{-s_{i}}{\delta_{i}}:\delta_{i} > 0\right\}, \max_{i}\left\{\frac{1-s_{i}}{\delta_{i}}:\delta_{i} < 0\right\}\right\}$$
$$\overline{\lambda} = \min\left\{\min_{i}\left\{\frac{1-s_{i}}{\delta_{i}}:\delta_{i} > 0\right\}, \min_{i}\left\{\frac{-s_{i}}{\delta_{i}}:\delta_{i} < 0\right\}\right\}$$

Next, for a dataset  $\mathbf{s} = (s_t)_{t \leq T} \in \Omega_{\text{GARP}}$ , define:

$$\ell(\mathbf{s}, t, \delta) = \operatorname{int} \left( \left\{ \lambda \in \Lambda(s_t, \delta) : (s_t + \lambda \delta, s_{-t}) \in \Omega_{\text{GARP}} \right\} \right),$$

where int denotes the interior. The set  $\ell(\mathbf{s}, t, \delta)$  equals the interior of the set of values  $\lambda$  such that (i) the share vector  $s_t + \lambda \delta$  is still in  $\Delta^k$  and (ii) the data set  $(s_t + \lambda \delta, s_{-t})$  satisfies GARP. Given the convexity of  $\mathcal{P}(s_{-t})$ , the set  $\ell(\mathbf{s}, t, \delta)$  is an open interval. Its endpoints can be found by binary search as shown in Algorithm 2. The algorithm finds (exponentially) better and better bounds on the upper and lower bound of this interval.

**Algorithm 2** Compute  $\ell(\mathbf{s}, t, \delta) = (a, b)$  up to an error  $\varepsilon$ 

```
Require: \mathbf{s} = (s_v)_{v \leq T} \in \Omega_{\text{GARP}}, t \leq T, \delta \in \partial D, \varepsilon > 0
   1: Compute \Lambda(s_t, \delta) = [\underline{\lambda}, \overline{\lambda}].
   2: b \leftarrow \lambda.
   3: b \leftarrow 0
  4: while (\overline{b} - \underline{b}) \ge \varepsilon/2 do
             b \leftarrow \frac{(\overline{b} + \underline{b})}{2}
   5:
             if (s_t + b\delta, s_{-t}) \in \Omega_{\text{GARP}} then
   6:
                  b \leftarrow b
   7:
              else
   8:
                   \overline{b} \leftarrow b.
   9:
              end if
 10:
 11: end while
 12: \overline{a} \leftarrow 0
 13: \underline{a} \leftarrow \underline{\lambda}
14: while (\overline{a} - \underline{a}) \ge \varepsilon/2 do
             a \leftarrow \frac{(\overline{a}+\underline{a})}{2}
if (s_t + a\delta, s_{-t}) \in \Omega_{\text{GARP}} then
 15:
 16:
                  \overline{a} \leftarrow a
 17:
 18:
              else
 19:
                   a \leftarrow a
 20:
              end if
21: end while
22: return (\overline{a}, b)
```

Algorithm 3 provides the entire MCMC algorithm. The algorithm starts from a dataset  $\mathbf{s}^0 = (s_t^0)_{t \leq T} \in \Omega_{\text{GARP}}$ . For this, we could, for example, use the dataset that gives equal shares to all goods over all observations. This data set is consistent with the optimization of a Cobb-Douglass utility function  $u(q_0, \ldots, q_k) = \prod_{i=0}^k q_i$ , so it satisfies GARP.

Then, for each iteration of the Markov chain we use the current dataset  $\mathbf{s}^n = (s_t^n)_{t \leq T}$ to produce a new dataset  $\mathbf{s}^{n+1}$ . This new dataset is generated by drawing an observation  $t \in \{1, \ldots, T\}$  at random (step 3) and using a H&R step to update  $s_t^n$  to a new share vector  $s_t^{n+1} \in \mathcal{P}(s_{-t}^n)$ . The H&R step is done by drawing a random direction  $\delta$  uniformly from  $\partial D$  (step 4); compute the set  $\ell(\mathbf{s}^n, t, \delta)$  (step 5) and draw a value  $\lambda$  uniformly from  $\ell(\mathbf{s}^n, t, \delta)$  (step 6). The new value  $s_t^{n+1}$  is given by  $s_t^n + \lambda \delta$ .

The fact that  $\lambda \in \ell(\mathbf{s}^n, t, \delta)$  guarantees that every newly generated data set remains in  $\Omega_{\text{GARP}}$ . As usual with a Gibbs sampler we can easily replace step 3 by choosing a random permutation  $\sigma : T \to T$  and then use the H&R to sequentially update all shares  $s_{\sigma(1)}^n, s_{\sigma(2)}^n, \ldots, s_{\sigma(T)}^n$ .

**Algorithm 3** Generate a Markov chain of data sets  $\mathbf{s}^n \in \Omega_{\text{GARP}}$ **Require:**  $\mathbf{s}^0 = (s^0_t)_{t \leq T} \in \Omega_{\text{GARP}}, M \in \mathbb{N}$ 1: Initialize  $n \leftarrow 0$ 2: while  $n \leq M$  do Randomly pick an observation  $t \in \{1, \ldots, T\}$ . 3:Draw a direction  $\delta$  uniformly from  $\partial D$  using Algorithm 1. 4: Compute  $(a, b) = \ell(\mathbf{s}^n, t, \delta)$  using Algorithm 2. 5: Draw  $\lambda$  uniformly from (a, b). 6: $s_t^{n+1} \leftarrow s_t^n + \lambda \, \delta$ 7:  $\begin{array}{c} \mathbf{for} \ v \neq t \ \mathbf{do} \\ s_v^{n+1} \leftarrow s_v^n \end{array}$ 8: 9: end for 10:  $n \leftarrow n+1$ 11: 12: end while

**Illustration** As an illustration, consider the following 2 goods, 2 observation example:

$$p_1 = (1, 1.5), p_2 = (1.5, 1).$$

It can be verified that the shares  $(s_t)_{t\leq 2}$  satisfy GARP if either  $s_{1,0} \geq 0.4$  or  $s_{2,0} \leq 0.6$ . In other words, we have a GARP violation if both  $s_{1,0} < 0.4$  and  $s_{2,0} > 0.6$ . Figure 1 shows randomly generated GARP consistent shares from this example using the algorithm given



Figure 1: Simulation results for a two goods two observations example

above. In particular, we use the random permutation version of the MCMC and we use a burn-in chain of 100 iterations of the Markov Chain. After this, we keep every 100th iteration. The top left picture shows the first 10 generated samples, the top right the first 100. The bottom left the first 250 and the bottom right the first 500. In this two goods, two observations example, our MCMC algorithm coincides with a standard Gibbs sampler over the GARP consistent region.

Theoretical properties of the MCMC In order to state the theoretical properties of our MCMC. We need to introduce some additional notation. We denote by  $P(\mathbf{s}, A)$  the transition kernel of the Markov Chain in Algorithm 3. In words  $P(\mathbf{s}, A)$  gives the one step probability of going from a dataset  $\mathbf{s} \in \Omega_{GARP}$  to a dataset in the measurable set  $A \subseteq \Omega_{GARP}$  in the next step. We define  $P^2(\mathbf{s}, A)$  by:

$$\int_{\Omega_{\rm GARP}} P(\mathbf{s}, d\mathbf{s}') P(\mathbf{s}', A)$$

as being the probability of going from  $\mathbf{s}$  to A in two iterations of the algorithm. Inductively, we define,

$$P^{n}(\mathbf{s}, A) = \int_{\Omega_{\text{GARP}}} P(\mathbf{s}, d\mathbf{s}') P^{n-1}(\mathbf{s}', A),$$

as the probability of going from  $\mathbf{s}$  to A in n iterations.

In order to show that the algorithm works, one needs to show two things. First, it is necessary to show that the uniform distribution  $\nu$  on  $\Omega_{\text{GARP}}$  is an invariant distribution for the Markov Chain. This means that:

$$\nu P = \nu$$

or equivalently,

$$\int_{\Omega_{\rm GARP}} \nu(d{\bf s}) P(s,A) = \nu(A)$$

for all measurable subsets A of  $\Omega_{\text{GARP}}$ . Next, one needs to show that at every starting point  $\mathbf{s} \in \Omega_{\text{GARP}}$ , the law of the MCMC sequence converges to this invariant distribution.<sup>8</sup>

$$||P^{n}(\mathbf{s},.) - \nu(.)|| \rightarrow^{n} 0 \qquad \forall \mathbf{s} \in \Omega_{\text{GARP}}.$$

The first invariance part is demonstrated by showing that the MCMC is reversible with respect to the uniform distribution  $\nu$  on  $\Omega_{\text{GARP}}$ .<sup>9</sup>

**Theorem 1.** The MCMC of Algorithm 3 is reversible, i.e. for all measurable subsets A and B of  $\Omega_{GARP}$ :

$$\int_{A} \nu(d\mathbf{s}) P(\mathbf{s}, B) = \int_{B} \nu(d\mathbf{s}) P(\mathbf{s}, A).$$

The proof of Theorem 1 is given in Appendix B. For the second part, we show an even stronger condition, namely uniform geometric ergodicity. This means that the convergence of the law of the MCMC converges to the invariant distribution at an exponential rate, uniformly over all starting positions  $\mathbf{s} \in \Omega_{\text{GARP}}$ .

**Theorem 2.** The MCMC of Algorithm 3 is uniformly geometric ergodic, i.e. there exists

<sup>&</sup>lt;sup>8</sup>Here, the norm is the total variation norm.

<sup>&</sup>lt;sup>9</sup>In order to see that reversibility implies invariance, notice that if P is reversible with respect to  $\nu$ , then for all measurable subsets B of  $\Omega_{\text{GARP}}$ :  $\int_{\Omega_{\text{GARP}}} \nu(d\mathbf{s}) P(\mathbf{s}, B) = \int_B \nu(d\mathbf{s}) P(\mathbf{s}, \Omega_{\text{GARP}}) = \int_B \nu(d\mathbf{s}) = \nu(B)$ .

a number M and a number  $r \in [0, 1]$  such that for all  $\mathbf{s} \in \Omega_{GARP}$ :

$$\|P^n(\mathbf{s},.) - \nu(.)\| \le Mr^n.$$

The proof of Theorem 2 can be found in Appendix C.

### 4 Other revealed preference tests

In this section, we show how the MCMC procedure can be extended to other revealed preference tests.

**WARP** A dataset  $\mathbf{s} = (s_t)_{t \leq T}$  is said to be consistent with the Weak Axiom of Revealed Preference or WARP if for all observations t, v:

$$t R v$$
 and  $v R t$  implies  $q_t = q_v$ .

WARP requires that the revealed preference relation is asymmetric (Samuelson, 1938). Let  $\Omega_{\text{WARP}}$  be the set of all WARP consistent datasets. Similarly to the analysis for GARP, we can consider the set:

$$\mathcal{P}_{\text{WARP}}(s_{-t}) = \{ \tilde{s}_t \in \Delta^k : (\tilde{s}_t, s_{-t}) \in \Omega_{\text{WARP}} \}.$$

Similarly to Lemma 1, one can show that this set is convex. The MCMC algorithm that we propose to sample uniformly from  $\Omega_{\text{WARP}}$  is almost identical to the algorithm for GARP with the sole exception that the set  $\ell(\mathbf{s}, t, \delta)$  is now replaced by:

$$\ell_{\text{WARP}}(\mathbf{s}, t, \delta) = \text{int} \left( \{ \lambda \in \Lambda(s_t, \delta) : (s_t + \lambda \delta, s_{-t}) \in \Omega_{\text{WARP}} \} \right).$$

The theoretical properties of this MCMC are identical to the one for GARP (the proof follows almost identical lines).

**HARP** A dataset  $\mathbf{s} = (s_t)_{t \leq T}$  is consistent with the Homothetic Axiom of Revealed Preference (HARP) if for all sequences  $t_1, \ldots, t_M$  of distinct observations:

$$(p_{t_1} \cdot q_{t_2})(p_{t_2} \cdot q_{t_3}) \dots (p_{t_M} \cdot q_{t_1}) \ge 1.$$

Varian (1983) showed that a dataset satisfies HARP if and only if it can be rationalised by a homothetic utility function. Let  $\Omega_{\text{HARP}}$  be the collection of datasets  $\mathbf{s} = (s_t)_{t \leq T}$ that satisfy HARP. Interestingly, one can demonstrate that the set  $\Omega_{\text{HARP}}$  is a convex set. Appendix E contains the proof.

#### **Lemma 2.** The set $\Omega_{HARP}$ is convex.

Given Lemma 2, we can use a standard H&R algorithm to sample from this dataset. For a set of directions  $(\delta_t)_{t\leq T}$  (i.e. for all  $t, \delta_t \in \partial D$ ), and a dataset  $\mathbf{s} \in \Omega_{\text{HARP}}$ , let:

$$\ell_{\text{HARP}}(\mathbf{s}, (\delta_t)_{t \le T}) = \text{int}\left(\{\lambda \in \bigcap_{t \le T} \Lambda(s_t, \delta_t) : (s_t + \lambda \delta_t)_{t \le T} \in \Omega_{\text{HARP}}\}\right)$$

The MCMC procedure to generate random datasets uniformly from  $\Omega_{\text{HARP}}$  is given by Algorithm 4

Algorithm 4 Generate a Markov chain of data sets  $D \in \Omega_{\text{HARP}}$ Require:  $\mathbf{s}^0 = (s_t^0)_{t \leq T} \in \Omega_{HARP}, M \in \mathbb{N}$ 1: Initialize  $n \leftarrow 0$ 2: while  $n \leq M$  do 3: Draw a set of directions  $(\delta_t)_{t \leq T}$  each one uniformly from  $\partial D$  using Algorithm 1. 4: Compute  $(a, b) = \ell_{\text{HARP}}(\mathbf{s}^n, (\delta_t)_{t \leq N})$  using binary search 5: Draw  $\lambda$  uniformly from (a, b). 6: for all  $t \leq T$ :  $s_t^{n+1} \leftarrow s_t^n + \lambda \delta_t$ 7:  $n \leftarrow n+1$ 8: end while

The left panel in Figure 2 shows a run of the MCMC algorithm for the same example as for Figure 1, but now for HARP.

**Quasi-linear** A dataset  $\mathbf{s} = (s_t)_{t \leq T}$  is consistent with the revealed preference conditions for quasi-linear utility maximization (QLRP) if for all sequences  $t_1, \ldots, t_M$  of distinct observations:

$$p_{t_1} \cdot (q_{t_2} - q_{t_1}) + p_{t_2} \cdot (q_{t_3} - q_{t_2}) + \ldots + p_{t_M} \cdot (q_{t_1} - q_{t_M}) \ge 0.$$

Brown and Calsamiglia (2007) showed that QLRP is a necessary and sufficient condition for rationalisability by a quasi-linear utility function.

Let  $\Omega_{\text{QLRP}}$  contain all data sets that are consistent with QLRP. Similarly as for HARP, it can be shown that this set is convex. As such, we can sample from this space using a



Figure 2: Simulation results for a two goods two observations case. Left for HARP and right for QLRP.

similar H&R as for the HARP test. Except that now we define:

$$\ell_{QLRP}(\mathbf{s}, (\delta_t)_{t \le T}) = \operatorname{int}\left(\{\lambda \in \bigcap_{t \le T} \Lambda(s_t, \delta_t) : (s_t + \lambda \delta_t)_{t \le T} \in \Omega_{QLRP}\}\right)$$

The right panel of Figure 2 shows the output of the 2 goods, 2 observations example for this case.

**Efficiency levels** In many occasions, one might be interested in weakening the revealed preference tests by introducing a so called efficiency level. Given a data set  $\mathbf{q} = (q_t)_{t \leq T}$ , let  $R^e$  be defined as:

$$q_t R^e q_v \leftrightarrow e \ge p_t q_v,$$
  
$$q_t P^e q_v \leftrightarrow e > p_t q_v.$$

Then e-GARP is defined as imposing the GARP restriction on the relations  $R^e$  and  $P^e$  instead of using R and P (Varian, 1990). Any dataset is consistent with e-GARP for e = 0. On the other hand if e = 1 then e-GARP is equivalent to GARP. As such, the closer e to one, the closer the dataset is to satisfying GARP.

Similarly, we can define a e-WARP test. For the HARP, we can say that a data set satisfies e-HARP if

$$\frac{(p_{t_1} \cdot q_{t_2})}{e} \frac{(p_{t_2} \cdot q_{t_3})}{e} \dots \frac{(p_{t_M} \cdot q_{t_1})}{e} \ge 1.$$

and for QLRP, we can define e-QLRP as:

$$p_{t_1} \cdot (q_{t_2} - e \, q_{t_1}) + p_{t_2} \cdot (q_{t_3} - e \, q_{t_2}) + \ldots + p_{t_M} \cdot (q_{t_1} - e \, q_{t_M}) \ge 0.$$

Sampling from the sets  $\Omega_{e-\text{GARP}}$ ,  $\Omega_{e-\text{WARP}}$ ,  $\Omega_{e-\text{HARP}}$  and  $\Omega_{e-\text{QLRP}}$  can be done easily by suitably redefining the sets  $\ell_{\text{GARP}}(\mathbf{s}, t, \delta)$ ,  $\ell_{\text{WARP}}(\mathbf{s}, t, \delta)$ ,  $\ell_{\text{HARP}}(\mathbf{s}, (\delta_t)_{t \leq T})$  and  $\ell_{\text{QLRP}}(\mathbf{s}, (\delta_t)_{t \leq T})$  defined above.

#### 5 Illustration

When applying a revealed preference test, it is customary to compute its so called Bronars power as a measure of the stringency of the revealed preference test. The Bronars power equals one minus the probability that a randomly generated data set passes the revealed preference test. In practice, this is performed using a Monte Carlo procedure. In a first step one draws a large number of data sets uniformly from  $(\Delta^k)^T$ . Next, one computes the share of all these randomly generated data sets that violate the revealed preference test. This corresponds to a standard Monte Carlo method to compute the size of the complement of the region  $\Omega_{\text{GARP}}$ , e.g. the value of  $1 - \mu(\Omega_{\text{GARP}})$ .

As usual with Monte Carlo integration, the (relative) accuracy of this computation will depend both the number of randomly generated datasets and on the size of the target. If the revealed preference test is very strict, i.e.  $\mu(\Omega_{GARP})$  is very small, then only a very small number of randomly generated datasets will satisfy the revealed preference test. In this case, the (relative) accuracy of this estimate will be very unreliable unless one generates a very large number of random data sets. Doing this becomes computationally very demanding.

To make the discussion somewhat more concrete, assume that we draw a large number of random datasets  $\mathbf{s}_i$  and then computing the fraction of these draws for which  $\mathbf{s} \in \Omega_{\text{GARP}}$ . The law of large numbers guarantees that this converges to  $\mu(\Omega_{\text{GARP}})$ .

$$\hat{\mu}(\Omega_{\text{GARP}}) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}[\mathbf{s}_i \in \Omega_{\text{GARP}}] \approx \mu(\Omega_{\text{GARP}}).$$

If  $\mu(\Omega_{\text{GARP}})$  is close to zero, however, the relative error:

$$\frac{\hat{\mu}(\Omega_{\rm GARP}) - \mu(\Omega_{\rm GARP})}{\mu(\Omega_{\rm GARP})},$$

will be large unless one uses a huge number of random samples.

It is possible to alleviate this problem by sequentially sampling from smaller and smaller subsets.<sup>10</sup> Assume, for example, that we have a sequence of nested models  $\Omega_{\text{GARP}} = \Omega_0 \subseteq \Omega_1 \subseteq \Omega_2 \ldots \subseteq \Omega_n = (\Delta^k)^T$ . Then by Bayes rule, we have:

$$\mu(\Omega_{\text{GARP}}) = \mu(\Omega_0|\Omega_1)\,\mu(\Omega_1|\Omega_2)\ldots\mu(\Omega_{n-2}|\Omega_{n-1})\,\mu(\Omega_{n-1}|\Omega_n) = \prod_{k=0}^{n-1} \frac{\mu(\Omega_k)}{\mu(\Omega_{k+1})},$$

If we can sample from  $\Omega_1, \Omega_2, \ldots, \Omega_n$ , then we can approximate each term on the right hand side by:

$$\frac{1}{N}\sum_{i=1}^{N}\mathbb{1}[\mathbf{s}_{(k+1),i}\in\Omega_k]\approx\mu(\Omega_k|\Omega_{k+1})=\frac{\mu(\Omega_k)}{\mu(\Omega_{k+1})},$$

where the datasets  $\mathbf{s}_{(k+1),i}$  are random draws from  $\Omega_{k+1}$ . If the value of  $\mu(\Omega_k)$  is not too small compared to  $\mu(\Omega_{k+1})$ , then these estimates are much more reliable and the resulting final estimate of  $\mu(\Omega_{\text{GARP}})$  will also be more reliable. The same reasoning can also be applied to compute the relative size of two (non-nested test) say  $\Omega_1$  and  $\Omega_2$ . If both  $\mu(\Omega_1)$ and  $\mu(\Omega_2)$  are very small, then the ratio:

$$\frac{\sum_{i} \mathbb{1}[\mathbf{s}_{i} \in \Omega_{1}]}{\sum_{i} \mathbb{1}[\mathbf{s}_{i} \in \Omega_{2}]},$$

will give a very poor estimate of  $\mu(\Omega_1)/\mu(\Omega_2)$ , as both numerator and denominator will be very small. However, assume that there is a third revealed preference test  $\Omega_0$  that is weaker than both  $\Omega_1$  and  $\Omega_2$ , i.e.  $\Omega_1 \cup \Omega_2 \subseteq \Omega_0$ . In this case, we can use the following approximation:

$$\frac{\mu(\Omega_1)}{\mu(\Omega_2)} = \frac{\mu(\Omega_1|\Omega_0)}{\mu(\Omega_2|\Omega_0)} \approx \frac{\sum_{i=1}^N \mathbb{1}[\mathbf{s}_{0,i} \in \Omega_1]}{\sum_{i=1}^N \mathbb{1}[\mathbf{s}_{0,i} \in \Omega_2]},$$

where the  $\mathbf{s}_{0,i}$ 's are now uniform draws from  $\Omega_0$ . The ratio on the right hand side is now a ratio of two numbers which are bigger, so this Monte Carlo approximation will give a much better approximation for a given sample size N. Again, in order to perform this computation, it is necessary that we can draw random data sets from  $\Omega_0$ , which can be done using the MCMC procedure outlined above.

As an illustration, let's simulate a dataset on normalized prices  $(p_t)_{t\leq T}$  with 20 ob-

 $<sup>^{10}\</sup>mathrm{In}$  the literature on Monte Carlo this procedure is called splitting.

servations and 10 goods. Normalized prices are uniformly drawn from the interval [4,5]. A simple Monte Carlo estimate, based on a sample of 100,000 random data sets gives  $\mu(\Omega_{\text{GARP}}) \approx 0$ . In order to obtain a more accurate estimate, consider the following decomposition:

$$\mu(\Omega_{\text{GARP}}) = \mu(\Omega_{\text{GARP}} | \Omega_{e^0 \text{-}\text{GARP}}) \times \prod_{i=0}^{n-1} \mu(\Omega_{e^i \text{-}\text{GARP}} | \Omega_{e^{i+1} \text{-}\text{GARP}}) \times \mu(\Omega_{e^n \text{-}\text{GARP}})$$

where  $e^i$  uses values from {0.998, 0.996, 0.993, 0.991, 0.988, 0.985, 0.80, 0.97, 0.95, 0}.<sup>11</sup> For each conditional probability, we use the MCMC procedure to generate 1000 random datasets consistent with  $e^{i+1}$ -GARP and compute the fraction of these that pass  $e^i$ -GARP to get an estimate of  $\mu(\Omega_{e^i-\text{GARP}}|\Omega_{e^{i+1}-\text{GARP}})$ . Doing this gives a final estimate of  $\mu(\Omega_{\text{GARP}}) \approx$  $8.58 \times 10^{-7}$ . The successive values of the conditional probabilities are given in Table 1.

i	$e^i$	$\mu(\Omega_{e^{i-1}\mathrm{GARP}} \Omega_{e^{i}\mathrm{GARP}})$
9	0.0	0.931
8	0.95	0.330
7	0.97	0.193
6	0.98	0.238
5	0.985	0.281
4	0.988	0.199
3	0.991	0.312
2	0.993	0.126
1	0.996	0.216
0	0.998	0.128

Table 1: Estimation of conditional probabilities to compute  $\mu(\Omega_{GARP})$ 

As a second illustration, let's try to compare the area of HARP and QLRP. Consider a setting with 10 observations and 5 goods where prices are drawn from the interval [4, 5]. Computing the area of HARP and QLRP consistent datasets gives a Monte Carlo esitmate of zero for both (based on 100,000 random samples), so it is impossible to get a value for  $\mu(\Omega_{\text{HARP}})/\mu(\Omega_{\text{QLRP}})$ . To get a better estimate, consider the following decomposition:

<sup>&</sup>lt;sup>11</sup>These values were chosen such that the estimates of  $\mu(\Omega_{e^i \text{GARP}} | \Omega_{e^{i+1} \text{GARP}})$  were not too small.

$$\frac{\mu(\Omega_{\text{HARP}})}{\mu(\Omega_{\text{QLRP}})} = \frac{\mu(\Omega_{\text{HARP}} | \Omega_{e^0\text{-}\text{HARP}}) \times \mu(\Omega_{e^0\text{-}\text{HARP}} | \Omega_{e^1\text{-}\text{HARP}}) \times \mu(\Omega_{e^1\text{-}\text{HARP}} | \Omega_{e^1\text{-}\text{GARP}})}{\mu(\Omega_{\text{QLRP}} | \Omega_{e^0\text{-}\text{QLRP}}) \times \mu(\Omega_{e^0\text{-}\text{QLRP}} | \Omega_{e^1\text{-}\text{GARP}})}$$

where  $e^0 = 0.995$  and  $e^1 = 0.99$ . This gives the estimate:

$$\frac{\mu(\Omega_{\text{HARP}})}{\mu(\Omega_{\text{QLRP}})} \approx \frac{(1814/10000)}{(2427/10000)} \frac{(5879/10000)}{(8081/10000)} \frac{(464/10000)}{(641/10000)} = 0.3936$$

which shows that, for this example, the area for HARP is about 39% of the area for QLRP.

### References

- Afriat, S. N., 1967. The construction of utility functions from expenditure data. International Economic Review 8, 67–77.
- Bélisle, C. J., Romeijn, H. E., Smith, R. L., 1993. Hit-and-run algorithms for generating multivariate distributions. Mathematics of Operations Research 18, 255–266.
- Bronars, S. G., 1987. The power of nonparametric tests of preference maximization. Econometrica 55, 693–698.
- Brown, D. J., Calsamiglia, C., 2007. The nonparametric approach to applied welfare analysis. Economic Theory 31, 183–188.
- Chambers, C. P., Echenique, F., 2016. Revealed Preference Theory. Cambridge University Press.
- Chen, M.-H., Schmeiser, B., 1993. Perofmance of the Gibbs, Hit and Run and Metropolis samplers. Journal of Computational and Graphical Statistics 2, 251–272.
- Crawford, I., De Rock, B., 2014. Empirical revealed preference. Annual Review of Economics, 503–524.
- Demuynck, T., Hjertstrand, P., 2020. Paul Samuelson, Master of Modern Economics. Palgrave MacMillan, Ch. Samuelson's Approach to Revealed Preference Theory: Some Recent Advances.

- Heufer, J., 2013. Generating random optimising choices. Computational Economics 44, 295–305.
- Houthakker, H. S., 1950. Revealed preference and the utility function. Economica 17, 159–174.
- Meyn, S. P., Tweedie, R. L., 1993. Markov Chains and Stochastic Stability. Springer-Verlag.
- Muller, M. E., 1959. A note on a method for generating points uniformly on n-dimensional spheres. Communications of the ACM 2, 19–20.
- Nummelin, E., 1984. General Irreducible Markov Chains and Non-Negative Operators. Cambridge University Press.
- Samuelson, P. A., 1938. A note on the pure theory of consumer's behavior. Economica 5, 61–71.
- Szpilrajn, E., 1930. Sur l'extension de l'ordre partiel. Fundamentae Mathematicae 16, 386–389.
- Varian, H. R., 1982. The nonparametric approach to demand analysis. Econometrica 50, 945–974.
- Varian, H. R., 1983. Non-parametric tests of consumer behavior. The Review of Economic Studies 50, 99–110.
- Varian, H. R., 1990. Goodness-of-fit in optimizing models. Journal of Econometrics 46, 125–140.

### A Construction of the uniform measure on $\Omega_{GARP}$

In this appendix, we show how to construct the uniform measure  $\nu$  over the sets  $\Omega_{\text{GARP}}$ . Towards this end, notice that every share vector  $s_t$  lies in the k dimensional unit simplex  $\Delta^{k}$ .<sup>12</sup> Let  $H^k$  be the space of all (not necessarily non-negative) k + 1 dimensional vectors

<sup>&</sup>lt;sup>12</sup>The vectors  $s_t$  have k + 1 components but one dimension is lost due to the adding-up constraint  $\sum_{i=0}^{k} s_i = 1.$ 

whose coordinates add up to one.

$$H^{k} = \left\{ x \in \mathbb{R}^{k+1} : \sum_{i=0}^{k} x_{i} = 1 \right\},\$$

Next, let  $L^k$  be the space of all k + 1 dimensional vectors whose coordinates add up to zero.

$$L^{k} = \left\{ x \in \mathbb{R}^{k+1} : \sum_{i=0}^{k} x_{i} = 0 \right\}.$$

There is a natural bijection  $f_1: H^k \to L^k$  given by  $f_1(x) = y$ , where  $y_i = x_i - 1/k$  for all  $i = 0, \ldots k$ . The set  $L^k$  is an (k-1) dimensional hyperplane through the origin. As such, it is possible to find k orthogonal base vectors  $e^1, \ldots e^k$  in  $L^k$ .

Let  $\tilde{e}^1, \ldots, \tilde{e}^k$  be the standard unit vectors on  $\mathbb{R}^k$ .<sup>13</sup> Define  $f_2: L^k \to \mathbb{R}^k$  such that for all  $i = 1, \ldots k$ :  $f(e^i) = \tilde{e}^i$  and for all  $x \in L^k$ :

$$f_2(x) = \sum_{i=1}^k \beta^k \tilde{e}^k$$
, where,  $x = \sum_{i=1}^k \beta^k e^k$ .

Finally, define  $\iota = f_1 \circ f_2 : H^k \to \mathbb{R}^k$ . The function  $\iota$  is an bijective isometry from  $H^k$  to  $\mathbb{R}^k$ .<sup>14</sup>

Let us denote by  $(\mathbb{R}^k, \mathcal{B}^k, \lambda^k)$  the measure space on  $\mathbb{R}^k$ , where  $\mathcal{B}^k$  equals the Borel  $\sigma$ -algebra on  $\mathbb{R}^k$  and  $\lambda^k$  is the corresponding Lebesgue measure. Using the isometry  $\iota$ , construct a corresponding measure space  $(H^k, \mathcal{A}, \phi)$ , where:

$$\mathcal{A} = \{ A \subseteq H^k : \iota(A) \in \mathcal{B}^k \},\$$
  
$$\phi(A) = \lambda^k(\iota(A))$$

<sup>&</sup>lt;sup>13</sup>The vector  $\tilde{e}^i$  has values  $\tilde{e}^i_j = 0$  if  $i \neq j$  and  $\tilde{e}^i_j = 1$  if i = j. <sup>14</sup>An isometry is a function that preserves Euclidean distances in the sense that  $||x - y|| = ||\iota(x) - \iota(y)||$ for all  $x, y \in H^k$ .

Let us also define  $(\mathbf{H}, \mathcal{D}, \mu)$  to be the *T*-fold product of  $(H^k, \mathcal{A}, \phi)$ :

$$\mathbf{H} = \underbrace{H^k \times H^k \dots \times H^k}_{T \text{ times}},$$
$$\mathcal{D} = \underbrace{\mathcal{A} \otimes \dots \otimes \mathcal{A}}_{T \text{ times}},$$
$$\mu = \underbrace{\phi \times \phi \dots \times \phi}_{T \text{ times}}.$$

From now on, when we say that a set (or function) is measurable, we mean that it is  $\mathcal{D}$ -measurable. We define  $\nu$  to be the uniform measure on  $\Omega_{\text{GARP}}$ : for all measurable sets  $A \in \mathcal{D}$ :

$$\nu(A) = \frac{\mu(A \cap \Omega_{\text{GARP}})}{\mu(\Omega_{\text{GARP}})}.$$

Strict positivity of  $\mu(\Omega_{\text{GARP}})$  follows from Lemma 6 below.

#### **B** Reversibility

Before proving reversibility of the Markov Chain, we need several Lemmata.

**Lemma 3.** For all non-negative, bounded measurable functions  $g : \mathbf{H} \to \mathbb{R}_+$  that satisfy  $g(\mathbf{s}) = 0$  if  $\mathbf{s} \notin \Omega_{GARP}$ , we have:

$$\frac{1}{\mu(\Omega_{GARP})}\int_{\mathbf{H}}gd\mu=\int_{\Omega_{GARP}}gd\nu.$$

*Proof.* Assume first that g is a simple function, i.e.  $\sum_i \beta_i \mathbb{1}[\mathbf{s} \in A_i]$  for a finite collection of non-negative numbers  $\beta_i$  and measurable sets  $A_i$ . Notice that we can restrict  $A_i \subseteq \Omega_{\text{GARP}}$ 

for all i. Then:

$$\begin{split} \frac{1}{\mu(\Omega_{\text{GARP}})} \int_{\mathbf{H}} g d\mu &= \frac{1}{\mu(\Omega_{\text{GARP}})} \int_{\mathbf{H}} \sum_{i} \beta_{i} \mathbb{1}[\mathbf{s} \in A_{i}] \mu(d\mathbf{s}), \\ &= \sum_{i} \beta_{i} \frac{1}{\mu(\Omega_{\text{GARP}})} \int_{\mathbf{H}} \mathbb{1}[\mathbf{s} \in A_{i}] \mu(d\mathbf{s}), \\ &= \sum_{i} \beta_{i} \frac{\mu(A_{i})}{\mu(\Omega_{\text{GARP}})}, \\ &= \sum_{i} \beta_{i} \frac{\mu(A_{i})}{\mu(\Omega_{\text{GARP}})}, \\ &= \sum_{i} \beta_{i} \int_{\mathbf{H}} \mathbb{1}[\mathbf{s} \in A_{i}] \nu(d\mathbf{s}), \\ &= \int_{\Omega_{\text{GARP}}} \sum_{i} \beta_{i} \mathbb{1}[\mathbf{s} \in A_{i}] \nu(d\mathbf{s}), \\ &= \int_{\Omega_{\text{GARP}}} g d\nu. \end{split}$$

Next, assume that g is the pointwise limit of an increasing sequence of simple functions  $g_n \uparrow g$ . Notice that, as  $g(\mathbf{s}) = 0$  for  $\mathbf{s} \notin \Omega_{\text{GARP}}$ , we can restrict for all  $n, g_n(\mathbf{s}) = 0$  for  $\mathbf{s} \notin \Omega_{\text{GARP}}$ . By the Monotone convergence theorem:

$$\frac{1}{\mu(\Omega_{\text{GARP}})} \int_{\mathbf{H}} g d\mu = \frac{1}{\mu(\Omega_{\text{GARP}})} \int_{\mathbf{H}} \lim_{n \to \infty} g_n d\mu,$$
$$= \lim_{n \to \infty} \frac{1}{\mu(\Omega_{\text{GARP}})} \int_{\mathbf{H}} g_n d\mu,$$
$$= \lim_{n \to \infty} \int_{\Omega_{\text{GARP}}} g_n d\nu,$$
$$= \int_{\Omega_{\text{GARP}}} \lim_{n \to \infty} g_n d\nu,$$
$$= \int_{\Omega_{\text{GARP}}} g d\nu.$$

**Lemma 4.** Let  $(\delta_t)_{t \leq T}$  be a set of k + 1 dimensional vectors such that for all  $t \leq T$ :  $\sum_{i=0}^{k} \delta_{t,i} = 0$  and let  $g: H \to \mathbb{R}_+$  be non-negative, measurable function. Then:

$$\int_{H} g(\mathbf{s})\mu(d\mathbf{s}) = \int_{H} g(\mathbf{s}+\delta)\mu(d\mathbf{s}).$$

*Proof.* Define the function  $f : \mathbf{H} \to \mathbf{H}$  such that  $f(\mathbf{s}) = (s_t + \delta_t)_{t \leq T}$ . By a change of variables, we have that for any integrable function  $g : \mathbf{H} \to \mathbb{R}_+$ :

$$\int_{H} gd(f_*\mu) = \int_{H} (g \circ f) d\mu$$

Where the push forward  $f_*\mu$  is defined as:

$$(f_*\mu)(A) = \mu(f^{-1}(A)) \qquad \forall A \in \mathcal{D}.$$

In our case:

$$(f_*\mu)(A) = \mu(A - \delta) = \mu(A),$$

where  $A - \delta = \{(s_t - \delta_t)_{t \leq T} : (s_t)_{t \leq T} \in A\}$ . The last inequality follows from the fact that  $A - \delta$  is a simple translation of A, which preserves the Lebesgue measure. As such:

$$\int_{\mathbf{H}} g(\mathbf{s})\mu(d\mathbf{s}) = \int_{\mathbf{H}} g(s)(f_*\mu)(d\mathbf{s}) = \int_{\mathbf{H}} g(f(\mathbf{s}))\mu(d\mathbf{s}) = \int_{\mathbf{H}} g(\mathbf{s}+\delta)\mu(d\mathbf{s}).$$

**Lemma 5.** Let  $(\delta_t)_{t\leq T}$  be a set of (k+1) dimensional vectors such that for all  $t \leq T$  $\sum_{i=0}^{k} \delta_{t,i} = 0$ . Let  $g : \mathbf{H} \to \mathbf{H}$  be a non-negative, measurable function such that  $g(\mathbf{s}) = 0$ and  $g(\mathbf{s} + \delta) = 0$  for all  $\mathbf{s} \notin \Omega_{GARP}$ . Then:

$$\int_{\mathbf{H}} g(\mathbf{s})\nu(d\mathbf{s}) = \int_{\mathbf{H}} g(\mathbf{s}+\delta)\nu(d\mathbf{s}).$$

*Proof.* We have:

$$\begin{split} \int_{\mathbf{H}} g(\mathbf{s})\nu(d\mathbf{s}) &= \frac{1}{\mu(\Omega_{\text{GARP}})} \int_{\mathbf{H}} g(\mathbf{s})\mu(d\mathbf{s}), \\ &= \frac{1}{\mu(\Omega_{\text{GARP}})} \int_{\mathbf{H}} g(\mathbf{s}+\delta)\mu(d\mathbf{s}), \\ &= \int_{\mathbf{H}} g(\mathbf{s}+\delta)\nu(d\mathbf{s}). \end{split}$$

The first and last lines follow from Lemma 3. The second line follows from Lemma 4.  $\Box$ 

Let  $P(\mathbf{s}, A)$  be the transition kernel for the MCMC Algorithm 3, i.e.  $P(\mathbf{s}, A)$  gives the probability that the next state is in A given that the current sate equals  $\mathbf{s}$ . Invariance of

the uniform distribution can be established by showing that the Markov chain is reversible: i.e. for all measurable subsets A, B of  $\Omega_{\text{GARP}}$ :

$$\int_{A} \nu(d\mathbf{s}) P(\mathbf{s}, B) = \int_{B} \nu(d\mathbf{s}) P(\mathbf{s}, A)$$

In order to see that this is true, notice that for all measurable sets B and all datasets  $\mathbf{s} \in \Omega_{\text{GARP}}$ :

$$P(\mathbf{s}, B) = \frac{1}{T} \sum_{t \le T} \int_{\partial D} \left( \int_{-\infty}^{\infty} \frac{\mathbb{1}[(s_t + \lambda \delta, s_{-t}) \in B]}{|\ell(\mathbf{s}, t, \delta)|} d\lambda \right) G(d\delta).$$

The first summation encompasses the random draw over the set of observations  $\{1, \ldots, T\}$ . Then the first integration formalizes the draw of the direction  $\delta$  over  $\partial D$ . Here, G is the uniform measure over this set. Finally, there is a random draw  $\lambda$  from the set  $\ell(\mathbf{s}, t, \delta)$ . This draw has density  $1/|\ell(\mathbf{s}, t, \delta)|$  with respect to the Lebesgue measure on  $\mathbb{R}$  when  $(s_t + \lambda \delta, s_{-t}) \in \Omega_{\text{GARP}}$  and has density zero outside of  $\Omega_{\text{GARP}}$ . However, in this latter case  $(s_t + \lambda \delta, s_{-t}) \notin B$  so the numerator then equals zero.

Using this we have that for all measurable sets  $A, B \subseteq \Omega_{GARP}$ :

$$\int_{A} \nu(d\mathbf{s}) P(\mathbf{s}, B) = \int_{A} \left( \frac{1}{T} \sum_{t \le T} \int_{\partial D} \left( \int_{-\infty}^{\infty} \frac{\mathbb{1}[(s_t + \lambda \delta, s_{-t}) \in B]}{|\ell(\mathbf{s}, t, \delta)|} d\lambda \right) G(d\delta) \right) \nu(d\mathbf{s}).$$

Using the independence of the random draw from  $\{1, \ldots, T\}$ ,  $\delta \in \partial D$ ,  $\lambda$  and  $\nu(d\mathbf{s})$  we can use Fubini's theorem to bring the integration on  $\nu(d\mathbf{s})$  inside:

$$\begin{split} \int_{A} \nu(d\mathbf{s}) P(\mathbf{s}, B) &= \frac{1}{T} \sum_{t \leq T} \int_{\partial D} \int_{-\infty}^{\infty} \left( \int_{A} \frac{\mathbb{1}[(s_{t} + \lambda \delta, s_{-t}) \in B]}{|\ell(\mathbf{s}, t, \delta)|} \nu(d\mathbf{s}) \right) d\lambda G(d\delta), \\ &= \frac{1}{T} \sum_{t \leq T} \int_{\partial D} \int_{-\infty}^{\infty} \underbrace{\int_{\mathbf{H}} \frac{\mathbb{1}[\mathbf{s} \in A] \mathbb{1}[(s_{t} + \lambda \delta, s_{-t}) \in B]}{|\ell(\mathbf{s}, t, \delta)|} \nu(d\mathbf{s}) d\lambda G(d\delta), \end{split}$$
(1)

Define the function  $g: \mathbf{H} \to \mathbb{R}_+$ , where:

$$g(\mathbf{s}) = \frac{\mathbb{1}[\mathbf{s} \in A]\mathbb{1}[(s_t + \lambda \delta, s_{-t}) \in B]}{|\ell(\mathbf{s}, t, \delta)|}$$

Notice that  $g(\mathbf{s}) = 0$  and  $g((s_t - \lambda \delta_t, s_{-t})) = 0$  when  $\mathbf{s} \notin \Omega_{\text{GARP}}$ . The latter one occurs as

in this case,

$$g((s_t - \lambda\delta, s_{-t})) = \frac{\mathbb{1}[(s_t - \lambda\delta, s_{-t}) \in A]\mathbb{1}[\mathbf{s} \in B]}{|\ell((s_t - \lambda\delta, s_t), t, \delta)|} = 0$$

as  $B \subseteq \Omega_{\text{GARP}}$ . As such, by Lemma 5:

$$\int_{\mathbf{H}} g(s)\nu(d\mathbf{s}) = \int_{\mathbf{H}} g((s_t - \lambda\delta, s_{-t}))\nu(d\mathbf{s}).$$

This gives:

$$\begin{aligned} \alpha &= \int_{\mathbf{H}} g(\mathbf{s})\nu(d\mathbf{s}), \\ &= \int_{\mathbf{H}} g((s_t - \lambda\delta, s_{-t}))\nu(d\mathbf{s}), \\ &= \int_{\mathbf{H}} \frac{\mathbb{1}[(s_t - \lambda\delta, s_{-t}) \in A]\mathbb{1}[\mathbf{s}_t \in B]}{|\ell((s_t - \lambda\delta, s_{-t}), t, \delta)|}\nu(d\mathbf{s}), \\ &= \int_B \frac{\mathbb{1}[(s_t - \lambda\delta, s_{-t}) \in A]}{|\ell(\mathbf{s}, t, \delta)|}\nu(d\mathbf{s}), \end{aligned}$$

Where we use the fact that  $|\ell((s_t - \lambda \delta, s_{-t}), t, \delta)| = |\ell(\mathbf{s}, t, \delta)|$ . If we plug this back in into (1), we obtain:

$$\begin{split} \int_{A} \nu(d\mathbf{s}) P(\mathbf{s}, B) &= \frac{1}{T} \sum_{t \leq T} \int_{\partial D} \int_{-\infty}^{\infty} \left( \int_{B} \frac{\mathbbm{1}[(s_{t} + \tilde{\lambda}\delta, s_{-t}) \in A]]}{|\ell(\mathbf{s}, t, \delta)|} \nu(d\mathbf{s}) \right) d\tilde{\lambda} G(d\delta), \\ &= \int_{B} \frac{1}{T} \left( \sum_{t \leq T} \int_{\partial D} \int_{-\infty}^{\infty} \frac{\mathbbm{1}[(s_{t} + \tilde{\lambda}\delta, s_{-t}) \in A]]}{|\ell(\mathbf{s}, t, \delta)|} d\tilde{\lambda} G(d\delta) \right) \nu(d\mathbf{s}), \\ &= \int_{B} \nu(d\mathbf{s}) P(\mathbf{s}, A). \end{split}$$

where we made the change of variables  $\tilde{\lambda} = -\lambda$ . This shows that the Markov chain is time reversible, so  $\nu$  is a stationary distribution.

## C Uniform geometric ergodicity

In this section, we show that the Markov chain is uniformly geometrically ergodic, which implies that the invariant distribution  $\nu$ , which coincides with the uniform distribution on  $\Omega_{\text{GARP}}$ , is the unique invariant distribution and that the Markov chain converges to this distribution exponentially fast.

As shown by Nummelin (1984, Theorem 5.16), it suffices to show the following two properties:

- 1. The set  $\Omega_{\text{GARP}}$  is small.
- 2. The transition kernel P is aperiodic,

We show aperiodicity of P in Theorem 4 below and we show smallness of  $\Omega_{\text{GARP}}$  in Theorem 3. For the this, we need to show the following *minorization* condition:

• We can find a constant  $\beta > 0$ , a number  $n \in \mathbb{N}$  and a probability measure  $\varphi$  on  $\Omega_{\text{GARP}}$  such that for all  $\mathbf{s} \in \Omega_{\text{GARP}}$  and all measurable subsets A of  $\Omega_{\text{GARP}}$ :

$$P^n(\mathbf{s}, A) \ge \beta \varphi(A).$$

In this case, we say that  $\Omega_{\text{GARP}}$  is  $(n, \beta, \varphi)$ -small.

Before we give the proofs of the two theorems, we first establish a useful lemma. Towards this end, let R be the revealed preference derived from a dataset  $\mathbf{s}$ . We say that the relation  $R^*$  extends the revealed preference relation on the dataset  $\mathbf{s}$  if for all observations  $t, v \leq T$ ,

$$t R v$$
 implies  $t R^* v$ .

A binary relation  $R^*$  (on  $\{1, \ldots, T\}$  is a *linear order* if it is transitive:  $t R^* v R^* w$  implies  $t R^* w$ , antisymmetric: if t R v and v R t then t = v and connected: either  $t R^* v$  or  $v R^* t$ .

**Lemma 6.** There exists a data set  $\mathbf{s} = (s_t)_{t \leq T}$  in the interior of  $\Omega_{GARP}$  and a linear order  $R^*$  on  $\{1, \ldots, T\}$  such that  $R^*$  extends the revealed preference relation on  $\mathbf{s}$ .

Further, there exists a number  $\varepsilon > 0$  such that for all datasets  $\tilde{\mathbf{s}} = (\tilde{s}_t)_{t \leq T}$  with,

$$\tilde{s}_t \in B_{\varepsilon}(s_t) = \left\{ \tilde{s}_t \in H^k : \|\tilde{s}_t - s_t\| < \varepsilon \right\},\$$

we have that  $\tilde{\mathbf{s}} \in \Omega_{GARP}$  and  $R^*$  also extends the revealed preference relation on  $\tilde{\mathbf{s}}$ .

*Proof.* Let  $\mathbf{s} = (s_t)_{t \leq T}$  be the data sets generated by the maximization of the Cobb-Douglass utility function

$$u(q) = \prod_{i=0}^{k} q_i.$$

This gives a share vector that is equal for all goods: for all observations  $t \leq T$  and all goods  $0 \leq i \leq k$ 

$$s_{t,i} = \frac{1}{k+1}.$$

As **s** is derived from the maximisation of a utility function, it satisfies GARP, so  $\mathbf{s} \in \Omega_{\text{GARP}}$ . Define the corresponding bundles  $\mathbf{q} = (q_t)_{t \leq T}$  where  $q_{t,i} = s_{t,i}/p_{t,i}$ . Let us first show that the revealed preference relation R on **s** is acyclic. Towards a contradiction, assume that there is a sequence  $t_1, \ldots, t_M$  such that:

$$t_1 R t_2 \dots R t_M R t_1.$$

As  $\mathbf{q} = (q_t)_{t \leq T}$  satisfies GARP it must be that:

$$1 = p_{t_1} \cdot q_{t_2}, \dots, 1 = p_{t_M} \cdot q_{t_1}.$$

Transforming this into shares and noticing that  $s_{t_j,i} = \frac{1}{k+1}$  for all good *i* and observation  $t_j$  in the sequence gives:

$$k + 1 = \sum_{i=0}^{k} \frac{p_{t_{1},i}}{p_{t_{2},i}},$$
  

$$k + 1 = \sum_{i=0}^{k} \frac{p_{t_{2},i}}{p_{t_{3},i}},$$
  

$$\dots,$$
  

$$k + 1 = \sum_{i=0}^{k} \frac{p_{t_{M},i}}{p_{t_{1},i}}.$$

Adding up these equalities and dividing by M(k+1) gives:

$$1 = \frac{1}{M(k+1)} \left( \sum_{i=0}^{k} \frac{p_{t_{1},i}}{p_{t_{2},i}} + \ldots + \sum_{i=0}^{k} \frac{p_{t_{M},i}}{p_{t_{1},i}} \right) > \left( \prod_{i=0}^{k} \frac{p_{t_{1},i}}{p_{t_{2},i}} \prod_{i=0}^{k} \frac{p_{t_{2},i}}{p_{t_{3},i}} \ldots \prod_{i=0}^{k} \frac{p_{t_{M},i}}{p_{t_{1},i}} \right)^{1/(M(k+1))} = 1,$$

a contradiction. Here, we used the fact that the average of non-negative numbers is always greater than the geometric mean if not all terms are equal (notice that we assumed price vectors to be distinct). This shows that R is acyclic. Therefore, we can extend it to a linear order  $R^*$  by Szpilrajn's lemma (Szpilrajn, 1930). Notice that for all observations  $t, v \leq T$ :

$$t R^* v \to 1 < p_v \cdot q_t = \sum_i p_{v,i} q_{t,i} = \sum_i p_{v,i} \frac{s_{t,i}}{p_{t,i}}$$

As all these inequalities are strict, we can find a  $\varepsilon > 0$  be such that for every  $t \leq T$ , every  $\tilde{s}_t \in B_{\varepsilon}(s_t) = \{\tilde{s}_t \in H^k : \|\tilde{s}_t - s_t\| < \varepsilon\}$ , and all observations  $t, v \leq T$ :

$$t R^* v \to 1 < \sum_i p_{v,i} \frac{\tilde{s}_{t,i}}{p_{t,i}} = \sum_i p_{v,i} \tilde{q}_{t,i} = p_v \cdot \tilde{q}_t, \tag{2}$$

where we defined the bundles  $\tilde{q}_{t,i} = \frac{\tilde{s}_{t,i}}{p_{t,i}}$ . Also, given that all share vectors  $s_t$  are in the interior of  $\Delta^k$ ,  $\varepsilon > 0$  can be chosen small enough such that all share vectors  $\tilde{s}_t$  are in  $\Delta^k$ . Let us show that any such dataset  $\tilde{\mathbf{s}} = (\tilde{s}_t)_{t \leq T}$  satisfies GARP and that  $R^*$  extends the revealed preference relation on  $\tilde{\mathbf{s}}$ . The second follows easily from the contrapositive of (2) and the fact that  $R^*$  is connected. For the first, let  $\tilde{R}, \tilde{P}$  be the revealed preference relations on  $\tilde{\mathbf{s}}$  and assume towards a contradiction that:

$$t_1 \,\tilde{R} \, t_2 \dots \tilde{R} \, t_M \,\tilde{P} \, t_1.$$

for some sequence  $t_1, \ldots, t_M$  of observations. This, in turn gives following inequalities:

$$1 \ge p_{t_1} \cdot \tilde{q}_{t_2}, 1 \ge p_{t_2} \cdot \tilde{q}_{t_3}, \dots, 1 \ge p_{t_{M-1}} \cdot \tilde{q}_{t_M}, 1 > p_{t_M} \cdot \tilde{q}_{t_1}.$$

But then, using again the contrapositive of (2) together with the fact that  $R^*$  is a linear order gives:

$$t_1 R^* t_2 R^* \dots t_M R^* t_1$$

which contradicts the fact that  $R^*$  is a linear order.

Now, we are ready to state and proof the smallness of  $\Omega_{\text{GARP}}$ . Let  $\mathbf{s} = (s_t)_{t \leq T}$  and  $B_{\varepsilon}(s_t)$  be as in Lemma 6. Consider the cylinder set  $B_{\varepsilon} = B_{\varepsilon}(s_1) \times B_{\varepsilon}(s_2) \times \ldots B_{\varepsilon}(s_T)$  and define for all measurable sets  $A \in \Omega_{\text{GARP}}$ :

$$\varphi(A) = \frac{\mu(A \cap B_{\varepsilon})}{\mu(B_{\varepsilon})}.$$

The measure  $\varphi$  coincides with the uniform measure on  $B_{\varepsilon}$ . The following shows that the set  $\Omega_{\text{GARP}}$  is small.

**Theorem 3.** There exists a constant  $\beta > 0$  such that for all  $\mathbf{s} \in \Omega_{GARP}$  and for all

measurable subsets A of  $\Omega_{GARP}$ :

$$P^T(\mathbf{s}, A) \ge \beta \varphi(A).$$

*Proof.* Notice that we can safely restrict A to be a subset of  $B_{\varepsilon}$ . Indeed, assume that the minorization constraint (i.e. the condition of the theorem) holds for all measurable sets  $A \subseteq B_{\varepsilon}$ . Then for  $\mathbf{s} \in \Omega_{\text{GARP}}$  and any measurable set C:

$$P^{T}(\mathbf{s}, C) = P^{T}(\mathbf{s}, C \setminus B_{\varepsilon}) + P^{T}(\mathbf{s}, C \cap B_{\varepsilon}),$$
  

$$\geq P^{T}(\mathbf{s}, C \cap B_{\varepsilon}),$$
  

$$\geq \beta \varphi(C \cap B_{\varepsilon}) = \beta \varphi(C).$$

so the minorization condition also holds for C. Let us first take the case where A takes the form of a cylinder set  $A = A_1 \times A_2 \times \ldots \times A_T$ . As we can restrict  $A \subseteq B_{\varepsilon}$ , we have that for all  $t \leq T$ ,  $A_t \subseteq B_{\varepsilon}(s_t)$ .

Let  $R^*$  be the linear extension of the preference relation as in Lemma 6. Consider the following permutation  $\sigma: T \to T$ :

- $\sigma(1) = t$  if t is top ranked according to  $R^*$ .
- $\sigma(2) = v$  if v is second highest ranked according to  $R^*$ ,
- ...
- $\sigma(k) = \ell$  if  $\ell$  is k-th ranked according to  $R^*$ ,
- . . .
- $\sigma(T) = w$  if w is bottom ranked according to  $R^*$ .

In other words:

$$\sigma(1) R^* \sigma(2) \dots \sigma(T-1) R^* \sigma(T).$$

Now, take any arbitrary dataset  $\mathbf{s} = (s_t)_{t \leq T} \in \Omega_{\text{GARP}}$ . Consider T iterations of Algorithm 3 where for the subsequent iterations, in steps 3, we obtain the following sequence of choices:

$$\sigma(1), \sigma(2), \ldots, \sigma(T)$$

from  $\{1, \ldots, T\}$  (the probability that this occurs is given by  $\left(\frac{1}{T}\right)^T$ ). Also assume that over these T iterations, the shares:

$$S_{\sigma(1)},\ldots,S_{\sigma(t)},\ldots,S_{\sigma(T)},$$

are sequentially updated to shares:

$$\tilde{s}_{\sigma(1)},\ldots,\tilde{s}_{\sigma(t)},\ldots,\tilde{s}_{\sigma(T)},$$

that belong (respectively) to the sets:

$$A_{\sigma(1)},\ldots,A_{\sigma(t)},\ldots,A_{\sigma(T)}.$$

Let us first show that this run of the algorithm is valid. In particular, let us show that at each iteration, replacing  $s_{\sigma(t)}$  by  $\tilde{s}_{\sigma(t)}$  does not create a GARP-violation. Towards a contradiction, assume that at iteration t, we introduce the share vector  $\tilde{s}_{\sigma(t)} \in A_{\sigma(t)}$  and

$$(\tilde{s}_{\sigma(1)},\ldots,\tilde{s}_{\sigma(t-1)},\tilde{s}_{\sigma(t)},s_{\sigma(t+1)},\ldots,s_{\sigma(T)}),$$

forms a GARP violation. Let

$$(\tilde{q}_{\sigma(1)},\ldots,\tilde{q}_{\sigma(t-1)},\tilde{q}_{\sigma(t)},q_{\sigma(t+1)},\ldots,q_{\sigma(T)}),$$

be the associated bundles for this GARP violation.

Notice that, from Lemma 6, both sub-datasets  $(\tilde{q}_{\sigma(1)}, \ldots, \tilde{q}_{\sigma(t)})$  and  $(q_{\sigma(t+1)}, \ldots, q_{\sigma(T)})$ satisfy GARP. As such, in order for there to be a GARP violation, there must be at least one observation  $\sigma(v)$  with v > t that is revealed preferred to the an observation  $\sigma(w)$  with  $w \leq t$ , i.e.:

$$1 \ge p_{\sigma(v)} \cdot \tilde{q}_{\sigma(w)}.$$

Now, given the definition of  $\sigma(.)$ , we have that  $\sigma(w) R^* \sigma(v)$ . As  $R^*$  extends the revealed preference relation on  $\tilde{\mathbf{q}}$ , we also have:

$$1 < p_{\sigma(v)} \cdot \tilde{q}_{\sigma(w)},$$

a contradiction.

Next, let us place a lower bound on the probability of such run. Using a result of Chen

and Schmeiser (1993, Lemma 3), we have that there is some constant K > 0 (independent from A and s) such that the probability of replacing  $s_{\sigma}(t)$  by some  $\tilde{s}_{\sigma}(t) \in A_{\sigma(t)}$  in the H&R part of the *t*-th iteration is bounded from below by:

$$K \ \lambda^k(\iota(A_{\sigma(t)})) = K \ \phi(A_{\sigma(t)}).$$

(see Appendix A for the definitions of  $\lambda^k$ ,  $\iota(.)$  and  $\phi$ ). This means that after T iterations of the MCMC in Algorithm 3, the probability of ending up with a dataset in  $A = A_1 \times \ldots A_T$  is bounded from below by:

$$\left(\frac{K}{T}\right)^T \prod_{t \le T} \phi(A_t) = \left(\frac{K}{T}\right)^T \mu(A) = \left(\frac{K}{T}\right)^T \mu(B_{\varepsilon})\varphi(A).$$

If we choose  $\beta = \left(\frac{K}{T}\right)^T \mu(B_{\varepsilon}) > 0$ , we see that:

$$P^T(\mathbf{s}, A) \ge \beta \varphi(A),$$

so the minorization condition is satisfied for all cylinder sets A. It is easy to see that the minorization constraint is closed under taking finite disjoint unions of sets, as both  $P^{T}(\mathbf{s}, .)$  and  $\varphi(.)$  are measures. Moreover the minorization constraint is also closed under taking increasing and decreasing sequences of measurable sets. As such, we can use the monotone class theorem in order to extend the minorization constraint to all measurable sets.  $\Box$ 

Finally, we show that P is aperiodic.

#### **Theorem 4.** The Markov transition kernel P is aperiodic.

*Proof.* The previous Theorem 3 shows that  $\Omega_{\text{GARP}}$  is a  $(T, \beta, \varphi)$  small set. This, in turn implies that P is  $\psi$ -irreducible. If P is not aperiodic, then by Meyn and Tweedie (1993, Theorem 5.4.4), there must be disjoint sets  $D_0, \ldots, D_{d-1}$  with d > 1 such that:

• if  $\mathbf{s} \in D_i$  then  $P(\mathbf{s}, D_{i+1}) = 1$ , for all  $i = 0, \dots d - 1 \pmod{d}$ ,

Let us show that this gives a contradiction. Let  $\mathbf{s} \in D_i$ , for some  $i \in \{0, \ldots, d-1\}$ . From Theorem 3, we know that:

$$P^T(\mathbf{s}, B_{\varepsilon}) > 0.$$

Also, for  $\mathbf{s} \in D_i$ :

$$P^T(\mathbf{s}, D_k) = 1$$
, where  $k = T + i \pmod{d}$ 

Let  $\varphi$  be defined as in Theorem 3. Let us first show that  $\varphi(D_k) = 1$ . If not, then

$$1 = \varphi(B_{\varepsilon}) \le \varphi(B_{\varepsilon} \setminus D_k) + \varphi(D_k) < \varphi(B_{\varepsilon} \setminus D_k) + 1$$

As such:

$$\varphi(B_{\varepsilon} \setminus D_k) > 0.$$

But then, by the fact that  $\Omega_{\text{GARP}}$  is  $(T, \beta, \varphi)$  small:

$$P^T(\mathbf{s}, B_{\varepsilon} \setminus D_k) > \beta \varphi(B_{\varepsilon} \setminus D_k) > 0,$$

Which implies that  $P^T(\mathbf{s}, D_k) < 1$ , a contradiction. By varying *i* we see that for all  $k = 0, \ldots d - 1, \varphi(D_k) = 1$ . As the sets  $D_0, \ldots, D_{d-1}$  are disjoint, we have:

$$1 \ge \varphi(\bigcup_{i=0}^{d-1} D_i) = \sum_{i=0}^{d-1} \varphi(D_i) = d,$$

which shows that d = 1, a contradiction.

### D Proof of Lemma 1

Let  $\mathbf{s} = (s_t)_{t \leq T} \in \Omega_{\text{GARP}}$  and let  $\mathbf{q} = (q_t)_{t \leq T}$  be the associated consumption bundles. Obviously,  $s_t \in \mathcal{P}(s_{-t})$ , so the set is non-empty. For convexity, observe that it is sufficient to show convexity of the set:

$$\mathcal{P}(q_{-t}) = \{ q_t \in \mathbb{R}^k_+ : p_t q_t = 1 \text{ and } (q_t, q_{-t}) \text{ satisfies GARP} \},\$$

as convexity of this set carries directly over to convexity of the set  $\mathcal{P}(s_{-t})$ . Towards this end, let  $q_t^0, q_t^1 \in \mathcal{P}(q_{-t})$ . Then both  $(q_t^0, q_{-t})$  and  $(q_t^1, q_{-t})$  satisfy GARP. Let  $\theta \in [0, 1]$  and define  $q_t^\theta = \theta q_t^0 + (1 - \theta) q_t^1$ . Towards a contradiction, assume that  $(q_t^\theta, q_{-t})$  violates GARP. Then, by definition there should be a direct revealed preference cycle containing at least one strict revealed preference comparison. Also, notice that this revealed preference cycle must contain the observation t as both datasets  $(q_t^0, q_{-t})$  and  $(q_t^1, q_{-t})$  satisfy GARP.

As GARP is violated, there are distinct observations  $t_1, \ldots, t_M \neq t$  such that

$$1 \ge p_t \cdot q_{t_1}, 1 \ge p_{t_1} \cdot q_{t_2}, \dots, 1 \ge p_{t_M} \cdot q_t^{\theta}.$$

where one of the inequalities is strict. The only place where the bundle  $q_t^{\theta}$  occurs is in the last inequality. Expanding this inequality gives:

$$1 \ge (>)p_{t_M} \cdot q_t^{\theta} = \theta p_{t_M} \cdot q_t^0 + (1-\theta)p_{t_M} \cdot q_t^1.$$

From this, it follows that either  $1 \ge p_{t_M} q_t^0$  (or  $1 > p_{t_M} q_t^0$  if the inequality is strict) or  $1 \ge p_{t_M} q_t^1$  (or  $1 > p_{t_M} q_t^1$  if the inequality is strict). If the first occurs, we have a GARP violation:

$$1 \ge p_t \cdot q_{t_1}, 1 \ge p_{t_1} \cdot q_{t_2}, \dots, 1 \ge p_{t_M} \cdot q_t^0.$$

where one of the inequalities is strict. If the second occurs, we have the GARP violation:

$$1 \ge p_t \cdot q_{t_1}, 1 \ge p_{t_1} \cdot q_{t_2}, \dots, 1 \ge p_{t_M} \cdot q_t^1.$$

where again at least one of the revealed preference comparisons is strict. So either  $q_t^0 \notin \mathcal{P}(q_{-t})$  or  $q_t^1 \notin \mathcal{P}(q_{-t})$  which gives the desired contradiction.

#### E Proof of Lemma 2

Let  $\mathbf{q}^0 = (q_t^0)_{t \leq T}, \mathbf{q}^1 = (q_t^1)_{t \leq T}$  satisfy HARP and let  $\theta \in [0, 1]$ . Define  $\mathbf{q}^\theta = (q_t^\theta)_{t \leq T}$  where for all  $t \leq T$ ,  $q_t^\theta = \theta q_t^0 + (1 - \theta) q_t^1$ . Towards a contradiction, assume that  $\mathbf{q}^\theta$  does not satisfy HARP. Then there are observations  $t_1, \ldots, t_M$  such that:

$$(p_{t_1} \cdot q_{t_2}^{\theta})(p_{t_2} \cdot q_{t_3}^{\theta}) \dots (p_{t_M} \cdot q_{t_1}^{\theta}) < 1.$$

Taking logs on both sides gives:

$$\ln(p_{t_1} \cdot q_{t_2}^{\theta}) + \dots + \ln(p_{t_M} \cdot q_{t_1}^{\theta}) < 0,$$
  

$$\leftrightarrow \ln(\theta p_{t_1} \cdot q_{t_2}^{0} + (1 - \theta) p_{t_2} \cdot q_{t_2}^{1}) + \dots + \ln(\theta p_{t_M} \cdot q_{t_1}^{0} + (1 - \theta) p_{t_M} \cdot q_{t_1}^{1}) < 0.$$

Then using concavity of the ln function, we obtain:

$$\theta \left( \ln(p_{t_1} \cdot q_{t_2}^0) + \ldots + \ln(p_{t_M} \cdot q_{t_1}^0) \right) + (1 - \theta) \left( \ln(p_{t_1} \cdot q_{t_2}^1) + \ldots + \ln(p_{t_M} \cdot q_{t_1}^1) \right) < 0$$

However, this means that either:

$$\ln(p_{t_1} \cdot q_{t_2}^0) + \ldots + \ln(p_{t_M} \cdot q_{t_1}^0) < 0,$$

or:

$$\ln(p_{t_1} \cdot q_{t_2}^1) + \ldots + \ln(p_{t_M} \cdot q_{t_1}^1) < 0.$$

As such, either  $\mathbf{q}^0$  or  $\mathbf{q}^1$  violates HARP, a contradiction.