AN EFFICIENT NONPARAMETRIC TEST OF THE COLLECTIVE HOUSEHOLD MODEL

RAHUL DEB†

ABSTRACT. The collective household model has emerged as one of the dominant models for studying household consumption behavior. This paper considers a special case of the collective model known as the *situation dependent dictatorship* model. This model is particularly suited to household scanner data where the bundle of goods purchased by the household is observed but the household member who made the purchasing decision is not. We show that the nonparametric test of this model is NP-Complete and provide an intuitive and computationally efficient approximate algorithm for implementing the test.

Keywords: Collective household models, intrahousehold allocation, revealed preference, nonparametric analysis, NP-Completeness, acyclic partition, directed graph, approximation algorithms.

†DEPARTMENT OF ECONOMICS, UNIVERSITY OF TORONTO, 150 ST. GEORGE ST., TORONTO, ONTARIO M5S 3G7
E-mail address: rahul.deb@utoronto.ca.
Date: August 27, 2010.
This paper supersedes an earlier version titled *Computational Inefficiency Of Nonparametric Tests For Consistency Of Data With Household Consumption*. I would like to thank Itzhak Gilboa, James Aspnes, Donald Brown and Maher Said for valuable discussions. Needless to say, all errors are mine.

1. Introduction

Consider the following problem. The bundles of goods purchased by a household over time and the associated prices $\{(p_i, x_i)\}_{i=1}^N$ are observed. However, the household member that made the purchasing decision at a particular time period is not observed. Moreover, it is not known how many decision makers are in household. We would like to infer from the data the number of decision makers and recover their preferences in a way that is consistent with the data. This problem was studied by Cherchye et al. (2007) who provide a simple necessary and sufficient condition for testing the data. This condition involves splitting the data into disjoint subsets, each of which satisfy the Generalized Axiom of Revealed Preference (henceforth GARP) when the entire data set may not. In this paper, we show that this test is a computationally hard problem and provide an efficient approximation algorithm for implementing the test.

There are many data sets now available which are good instances of the above problem. Household supermarket data is a nice example of a case where the bundle of goods purchased depends on the person within the household whose turn it is to do the chores. Household scanner panels are compiled by providing households with scanners so that they can track their purchases over a period of time (Nielsen Homescan and Information Resources Inc. Consumer Network are two major providers of such data). These data sets are collected over many years in different countries and include a variety of different products. They also have considerably more observations per household than traditional nationwide consumer panels which make computational issues quite relevant. Moreover, this data also has substantial relative price variation across time and hence can be tested in a meaningful way. Big retail stores also maintain similar scanner data. They track repeat purchases of a household by the use of loyalty cards. These panel data sets also contain many more observations per household than traditional consumer panels. There is an extensive and current marketing literature which utilizes such data sets (to cite a few significant papers, we mention Ainslie and Rossi (1998), Bucklin and Gupta (1999) and Swait and Andrews (2003)).

The above problem is a special case of the more general collective household model originally introduced by Chiappori (1988, 1992). This model was the first notable divergence from the unitary approach, which assumes that each household has a single utility function, that is, it acts as a single decision maker. The collective household model differs from the unitary model in the fact that it allows for (possibly diverging) individual preferences amongst different members of the household. It merely assumes that the household behaves Pareto optimally, with the weight of each member's utility function providing a measure of the distribution of power. This model has proved to be a critical tool for studying household consumption and has been widely utilized and tested in the empirical literature. [1]

1.1. *Discussion of Results and Related Literature*

As mentioned above, we show that splitting a data set into subsets that satisfy GARP is a NP-Complete problem. This means that there is no known polynomial (in the number of data points) time algorithm which can solve the problem unless P=NP. NP-Completeness is perhaps the most widely accepted notion of complexity. That said, it should be noted at the outset that it is a worst case criterion. It is conceivable that problems which are hard with respect to this measure have good expected complexity, that is, are computationally efficient on average. The relevance of the result in this paper is affirmed by recent empirical work in this area which suggests that the simplest partitioning approach would perhaps not scale well to larger data sets (Cherchye et al. 2008, 2009).

Dobell (1965) and Varian (1982) have pointed out that the revealed preference relation generates a directed graph with each observed price and consumption bundle as a vertex. There is a directed edge between two vertices if a bundle is revealed preferred to another. The acyclic partitioning problem for undirected graphs is a known NP-Complete problem. Unsurprisingly, the equivalent problem for directed graphs can also be shown to be NP-Complete (see Deb (2010) for the harder case of partitioning into two acyclic partitions). In light of this, the complexity result in this paper might seem trivial. A particular instance of partitioning consumption data into subsets that satisfy GARP is equivalent to the acyclic partitioning of the directed graph generated by using the revealed preference relation which is a NP-Complete problem.

However, showing the computational inefficiency of the nonparametric test requires us to go the other way. We show that given an arbitrary directed graph, it is always possible to construct a data set which generates the given graph through the revealed preference relation. In our opinion, this is a surprising result as one expects the revealed preference relation to impose some restriction on the set of underlying graphs. A worst case instance of the acyclic partitioning problem on

[1] A few influential papers that test the household model are Udry (1996), Dercon, S. and Krishnan, P. (2000) and Duflo and Udry (2004).

directed graphs need not arise from observed consumption data. The revealed preference relation is generated a consumer's choice from various budget sets. It is not obvious that bundles of goods and budget sets can be constructed which intersect and overlap in appropriate ways to generate any arbitrary directed graph. Moreover, an important consideration is that this construction must be done in polynomial time. We feel that this construction is one of the main insights of this paper.

In light of the complexity result, we provide an approximation algorithm to answer the following question: What is the minimum number of decision makers that could have generated the observed data? This involves finding the smallest number of disjoint subsets that the data set can be split into so that each subset satisfies GARP. The complexity result implies that this problem is NP-Hard and hence we design an efficient algorithm which can give us an approximate solution. We present such an algorithm which approximately solves the minimum acyclic partitioning problem on directed graphs (or equivalently the nonparametric test). This algorithm adapts the algorithm of Chen (2000) which solves the same problem for undirected graphs. Chen's result exploits a well known upper bound for the minimum number of acyclic partitions in undirected graphs. To the best of our knowledge such a bound does not exist for directed graphs. We theoretically establish such a bound and this allows us to adapt Chen's algorithm.

Since this paper, there have been other interesting efforts to provide computationally efficient solutions of related problems. In an insightful contribution, Cherchye et al. (2010b) provide a nonparametric characterization of the general collective model which they call the Collective Axiom of Revealed Preference (CARP). Talla Nobibon et al. (2010) adapt methods in this paper to show that testing CARP is NP-Complete. Cherhcye et al (2010a) have also developed Integer Programming techniques which address computational issues in testing General Equilibrium models.

This paper is organized as follows. Section 2 describes the model and sets up the problem. Section 3 presents the complexity result and its proof. Section 4 describes an efficient algorithm which computes an approximate solution to the problem and the algorithm is implemented on simulated data. Section 5 concludes.

2. The Model

We consider a K person household. There are L goods. An observed *data set* is a finite set of price, household consumption vectors $D = \{(p_i, x_i)\}_{i=1}^{N}$, where $(p_i, x_i) \in \mathbb{R}^L_{++} \times (\mathbb{R}^L_{+} \backslash \{0\})$ and

$1 \leq N < \infty$. It is assumed that at each observation, one of the K members of the household chooses the bundle of goods in accordance with their preferences. Each household member $k \in \{1, \ldots, K\}$ has a utility function given by

$$U^k : \mathbb{R}^L_+ \backslash \{0\} \to \mathbb{R}^L_+.$$

We define the Generalized Axiom of Revealed Preference (GARP) as in Varian (1982).

Definition 2.1 (GARP)**.** Given arbitrary data set $D = \{(p_i, x_i)\}_{i=1}^N$. For any two consumption bundles x_1 and x_2 we say $x_1 \succ_{R^0} x_2$ if $p_1' x_1 \geq p_1' x_2$. We say $x_1 \succ_P x_2$ if $p_1' x_1 > p_1' x_2$. Finally we say $x_1 \succ_R x_2$ if for some sequence of observations $(x_i, x_j, \ldots, x_m)$ we have $x_1 \succ_{R^0} x_i$, $x_i \succ_{R^0} x_j$, $\ldots, x_m \succ_{R^0} x_2$. In other words relation $\succ_R$ is the transitive closure of $\succ_{R_0}$. The data D satisfies ***GARP*** if

$$x_i \succ_R x_j \implies x_j \not\succ_P x_i \qquad \forall x_i, x_j$$

Varian (1982) showed that a data set is consistent with utility maximization if and only if it satisfies GARP. His theorem is presented below.

Theorem 1 (Varian)**.** *Let $D = \{(p_i, x_i)\}_{i=1}^N$ be an individual's price consumption data set. The following are equivalent*

(1) *Data set D is consistent with utility maximization. In other words, there exists a nonsatiated utility function U such that for each observation i*

$$U(x') \leq U(x_i),$$

$$\text{for all } x' \text{ satisfying } p_i x' \leq p_i x_i.$$

(2) *Data set D satisfies GARP.*

We now define the notion of consistency with K dictators which is the focus of this paper.

Definition 2.2 (Consistency with K Dictators)**.** Let $D = \{(p_i, x_i)\}_{i=1}^N$ be an observed household consumption data set. We say the data D is ***consistent with*** K ***dictators*** if there exist nonsatiated utility functions $U^1, U^2, \ldots, U^K$ such that for each observation i there is an individual $1 \leq k \leq K$ for whom

$$U^k(x') \leq U^k(x_i),$$

for all x' satisfying $p_i x' \leq p_i x_i$.

Cherchye et al. (2007) observe that a data set is consistent with K dictators if and only if the data set D can be split into K disjoint sets $\{D^1, \ldots, D^K\}$ such that each subset satisfies GARP. Then the consumption bundles in D^k can be thought of as the choices made by household member k which must be consistent with utility maximization and hence must satisfy GARP.

3. The Complexity Result

We now present and prove our main result which shows that partitioning a data set into disjoint subsets that satisfy GARP is an NP-Complete problem. As we mentioned above, we feel that the main insight of this section is not the complexity result itself but the proof which shows that the revealed preference relation imposes no restriction on the underlying graph it generates. In addition, we show that it is possible to construct the data set corresponding to an arbitrary directed graph in polynomial time which is critical.

Theorem 2. *Given a fixed integer* $1 < K \leq N$. *Testing the consistency of the data with K decision makers is NP Complete in the number of data points* N.

Proof. As mentioned previously, we will use the acyclic partitioning problem on directed graphs to show this result. The equivalent problem on undirected graphs is the partition into forests problem which is a known NP-Complete problem. Garey and Johnson (1979, pg 193) show that for $K \geq 3$ the problem of splitting an undirected graph into K trees is NP-Complete. They use the graph 3 coloring problem to show this result. The acyclic partitioning problem for directed graphs can be shown to be NP-Complete for $K \geq 3$ also using a reduction from graph 3 coloring. For $K = 2$, Wu et al. (1996) show that the problem of partitioning an undirected graph into $K = 2$ trees is NP Complete. They use the one-in-three 3Sat problem (pg 259 in Garey and Johnson 1979) for the reduction. Partitioning a directed graph into $K = 2$ acyclic components has been shown by Deb (2010) to be NP-Complete by using a reduction from the not-all-equal 3Sat problem (pg 259 in Garey and Johnson 1979).

We will now proceed by mapping every instance of the acyclic partitioning problem on directed graphs to a particular instance of the data consistency problem. Consider a directed graph $G(V, E)$ and a given fixed positive integer $1 < K \leq |V|$, where V is the set of vertices and E is the set of edges. We say there is a directed edge from vertex i to j if $(i,j) \in E$. Finally we define the out degree of a vertex i as the number of outward edges from the vertex i. The acyclic K partitioning problem is then defined as follows.

Given an arbitrary integer $K > 1$, can the nodes of the graph be partitioned into $s \leq K$ disjoint sets $V_1, V_2, \ldots, V_s$ in such a way that for each V_j $(1 \leq j \leq s)$ the subgraph $G_j(V_j, E_j)$ induced by V_j contains no cycles? Such a partition is called an acyclic K partition.

We start with an arbitrary directed graph $G(V,E)$ where $|V| = N$. We now construct an appropriate data set $\{(p_i, x_i)\}_{i=1}^N$ (in polynomial time) such that the graph induced by the relation $\succ_P$ on the data is the same as G. We proceed as follows.

Let us assume that the number of goods in each bundle is N^2. Each good is labeled by two numbers mn where $1 \leq m,n \leq N$. In a slight abuse of notation, we represent the mn^{th} good in the i^{th} observation by x_i^{mn}, and the price of the mn^{th} good in the i^{th} observation by p_i^{mn}, where $1 \leq i,m,n \leq N$.

We now construct the data set $D = \{(p_i, x_i)\}_{i=1}^N$, where price consumption bundle i corresponds to vertex $i \in V$, as follows

$$
\begin{array}{llll}
p_i^{ii} = 1 & x_i^{ii} = 1 & & \\
p_i^{ij} = 1 & x_i^{ij} = 1 & & \text{if } j \neq i \text{ and } (i,j) \in E \\
p_i^{ji} = 0 & x_i^{ji} = 0 & & \text{if } j \neq i \text{ and } (j,i) \in E \\
p_i^{ij} = 1 & x_i^{ij} = 0 & & \text{if } j \neq i \text{ and } (i,j) \notin E \\
p_i^{ji} = 0 & x_i^{ji} = \text{out degree of } (j) + 2 & & \text{if } j \neq i \text{ and } (j,i) \notin E \\
p_i^{kl} = 0 & x_i^{kl} = 0 & & \text{if } k,l \neq i
\end{array} \tag{1}
$$

Clearly this data set can be constructed in polynomial time. The intuition behind the construction is as follows. Each observation i corresponds to a vertex. Each good corresponds to an edge of the graph, except goods of the form ii which ensure each observation reflects positive wealth. For any two observations i and j we select appropriate prices and consumption bundles for the goods ij and ji (the remaining goods not coming into play) so that the relation $\succ_P$ reflects whether there is an edge (i,j) or (j,i) in the graph G.

We now check to see if the graph induced by the relation $\succ_P$ on the above data set is the same as G. For an arbitrary $i \neq j$ and $i, j \in V$, if we have $(i, j) \in E$ then

$$\begin{aligned}
&p_i' x_i = \text{out degree of } (i) + 1 \qquad \text{[which is at least 1]} \\
&p_i' x_j = 0 \\
&\implies p_i' x_i > p_i' x_j \\
&\implies x_i \succ_P x_j
\end{aligned}$$

Similarly if we have $(i, j) \notin E$ then

$$\begin{aligned}
&p_i' x_i = \text{out degree of } (i) + 1 \\
&p_i' x_j = \text{out degree of } (i) + 2 \\
&\implies p_i' x_i < p_i' x_j \\
&\implies x_i \not\succ_P x_j
\end{aligned}$$

In the data set defined above, we have $x_i \succ_P x_j$ (equivalently $x_i \succ_{R^0} x_j$) if, and only if there is an edge from vertex i to vertex j in G (or $(i, j) \in E$). Hence the graph induced by the relation $\succ_P$ on the above data set is the same as G. Moreover, any subset $D' \subseteq D$ of the data set satisfies GARP if, and only if, the subgraph induced by the vertices corresponding to the data points in D' is acyclic. Hence, we can solve the acyclic K partitioning problem for arbitrary graph G and fixed integer K, if, and only if, we can check the consistency of the data set D with K decision makers.

The proof is not complete because our model does not allow data with zero prices or bundles where nothing is consumed. To solve this problem we can replace every instance of 0 by a very small $\epsilon < 1$ in equation (1). We can once again verify the inequalities for each i. If $i \neq j$, $i, j \in V$ and $(i, j) \in E$

$$\begin{aligned}
&p_i' x_i > \text{out degree of} (i) + 1 \\
&p_i' x_j < N^2 (N + 2)\epsilon \qquad \text{[Because each vertex can have out degree at most } N\text{]}
\end{aligned}$$

If $(i,j) \notin E$

$$p_i' x_i < \text{out degree of}(i) + 1 + N^2(N+2)\epsilon$$

$$p_i' x_j > \text{out degree of}(i) + 2$$

For small enough $\epsilon < \frac{1}{N^2(N+2)}$ we will have $p_i'x_i > p_i'x_j$ when $(i,j) \in E$ and $p_i'x_i < p_i'x_j$ when $(i,j) \notin E$.

Hence, we can map (in polynomial time) any arbitrary graph G to a particular data set D such that for a fixed integer K we can solve the acyclic K partitioning problem for G if, and only if, the data set D is consistent with K decision makers. Thus testing an arbitrary data set for consistency with K decision makers is NP Complete. □

We would like to point out that the equivalence in the above proof can be established by assuming fewer than N^2 goods. However this results in considerable complication without further insight.

4. An Efficient Algorithm For Acyclic Partitioning of Directed Graphs

4.1. *The Algorithm*

In this section, we address the following question: What is the minimum K such that the observed data is consistent with K dictators? For simplicity, we make the following assumption. We assume that the observed data set $D = \{(p_i, x_i)\}_{i=1}^N$ contains no two observations $j \neq k$ such that $p_j'x_j = p_j'x_k$. This basically implies that there can be no two observations j, k such that at prices p_j, the bundles x_j and x_k cost the same.[2] In our opinion, this assumption is innocuous because it is extremely unlikely to encounter a violation of this assumption in consumption data which normally documents expenditures in dollars and cents. This assumption allows us to ignore indifferences or in other words we only need to consider the strict preference relation $\succ_P$. Hence, equivalently the problem we need to solve is the following. *Given a directed graph, what is the minimum K so that the vertices of the graph can be partitioned into K acyclic components?* This minimum K is referred to

[2] Equivalently it possible to redefine the setting in terms of the Strong Axiom of Revealed Preference (see Varian (1982)). But this will involve further notation which in our opinion is merely distracting.

as the arboricity of a graph. This is a NP-Hard problem and hence we now design an algorithm which approximately solves it.

The corresponding problem for undirected graphs has been studied in the theoretical computer science literature. That problem is also NP-Hard. There is a good upper bound for the arboricity of undirected graphs which is well known and Chen (2000) provides an efficient algorithm to solve the partitioning problem on undirected graphs with respect to this upper bound. To the best of our knowledge, the corresponding problem has not been solved for directed graph. We establish a similar upper bound for directed graphs and adapt the algorithm of Chen (2000) to solve the problem.

Recall, $G = (V, E)$ is a directed graph where V is the set of vertices and E is the set of edges. For any vertex $v \in V$, $G - v$ refers to the graph induced by the vertices $V \backslash \{v\}$. An *acyclic K partition* of G is a partition V_i, $1 \leq i \leq K$ of vertices V such that the subgraphs $G_i(V_i, E_i)$ induced by vertices V_i are acyclic. The *arboricity* $a(G)$ is the minimum number K for which G has an acyclic K partition. Arboricity is a measure of the denseness of a graph. As mentioned earlier, the problem of computing $a(G)$ for a graph is NP Hard. We first define a good upper bound for $a(G)$. A similar bound for undirected graphs can be found in Chartrand and Kronk (1969). We adapt their proofs to establish our bound.

For any vertex $v \in V$ we define $\delta(v) = \min\{\text{out degree of }(v), \text{in degree of }(v)\}$. We define $\Delta(G) = \min_{v \in V} \delta(v)$. A graph G is called a *critical graph* if $a(G - v) < a(G)$ for all $v \in G$ and a critical graph G with $a(G) = K$ is called a *K critical graph.* A K critical graph necessarily has $K \geq 2$, moreover, every graph G with $a(G) = K \geq 2$ has a subgraph G' which is K critical (consider the subgraph G' with the fewest vertices such that $a(G') = K$). We are now in a position to derive the bound for $a(G)$.

Lemma 1. *A K critical graph G must have $\Delta(G) \geq K - 1$.*

Proof. Suppose G contains a vertex v such that $\delta(v) \leq K - 2$. Since G is K critical it must be the case that $a(G - v) = K - 1$. Hence the vertices $V - v$ can be partitioned into $V_1, \ldots, V_{K-1}$ such that the graph induced by each V_i, $1 \leq i \leq K - 1$ is acyclic.

Because $\delta(v) \leq K - 2$, there must exist some V_i such that v either has only outward edges to or has only inwards edges from the vertices in V_i. Then $V_i \cup v$ must be acyclic and this contradicts the fact that $a(G) = K$ as we have found an acyclic $K - 1$ partition of G. □

We now derive the desired upper bound.

Lemma 2. *For arbitrary graph G it must be the case that $a(G) \leq \rho(G)$ where $\rho(G) = 1 + \max \Delta(G')$, where the maximum is taken over all subgraphs G' of G.*

Proof. The result is obvious if G is acyclic. Hence, let G be a graph with $a(g) = K \geq 2$. Let H be a subgraph of G such that H is K critical. We know that $\Delta(H) \leq \rho(G) - 1$.

By lemma 5.1 we know

$$\begin{aligned} & \Delta(H) \geq K - 1 \\ \implies \quad & \rho(G) - 1 \geq K - 1 = a(G) - 1 \\ \implies \quad & a(G) \leq \rho(G) \end{aligned}$$

□

This bound is "tight" for planar, outerplanar[3] and complete graphs. We show this by using similar known results on undirected graphs. For a vertex $v \in V$ in an undirected graph, we abuse notation and define $\delta(v) =$ degree of (v). The definition of $\Delta(G)$ remains unchanged but is now a minimum over the redefined δ. Chartrand and Kronk (1969) show that an upper bound for the arboricity of an undirected graph is then $a(G) \leq 1 + \lfloor \max \Delta(G')/2 \rfloor$ where the maximum is taken over all subgraphs G' of G. There is an intuitive similarity between this bound and the one established in lemma 5.2. Chartrand and Kronk (1969) show that this bound is tight, that is, $\rho(G) \leq 3$ for undirected planar graphs. This result stems from the observation that every planar graph contains a vertex of degree 5 or less and that every subgraph of a planar graph is planar. Using a similar argument, they also show that $\rho(G) \leq 2$ for undirected outerplanar graphs. It is straightforward to see that these results carry over verbatim to the case of directed planar (i.e. $\rho(G) \leq 3$) and outerplanar graphs (i.e. $\rho(G) \leq 2$). Moreover, the bound is trivially tight for

[3]In these graphs there is at most one directed edge between a pair of vertices.

complete directed graphs. Hence, $\rho(G)$ is a good upper bound for sparse graphs and graphs with high edge density.

While we present no theoretical results for the tightness of this bound on arbitrary graphs, it should be noted that the revealed preference relation induces sparse graphs for most household consumption data. In other words, while household data does violate GARP the number of such violations are not large (see for example Famulari 1995) and hence the induced graphs will have low arboricity. Moreover, household consumption data is most commonly tested for consistency with 2 decision makers. If the true underlying model consists of 2 decisions makers then the arboricity of the graph induced by the data will be 2 and hence the graph will be sparse. $\rho(G)$ should then be a good approximation of the arboricity and this is confirmed by the simulations in the next subsection. Finally, simulations also suggest that in general for arbitrary graphs, $\rho(G)$ is a good approximation of the arboricity.

We now define a simple and efficient algorithm to compute an acyclic $\rho(G)$ partition of an arbitrary directed graph G. This algorithm is an adaptation of the algorithm in Chen (2000) for undirected graphs.

Algorithm

Input: A directed graph $G = (V, E)$.

Output: An acyclic $\rho(G)$ partition of G.

(1) If $|V| = 1$, then output the unique vertex in G as the partition and halt.
(2) Find a vertex v such that $\delta(v) = \Delta(G)$.
(3) Recursively, call the algorithm on the graph $G - v$. Let $V_1, \ldots, V_k$ be the returned partition.
(4) Let $k' = 1 + \Delta(G) = 1 + \delta(v)$. Find the first set V_i among $V_1, \ldots, V_k$ such that graph induced by $V_i \cup \{v\}$ is acyclic. If such a set exists we assign v to set V_i. If not (then it must be the case that $k < k'$), add another set $\{v\}$ to the partition.

The above algorithm can be shown to be correct by induction on the number of vertices of G. When G has only one vertex the algorithm vacuously gives the correct output. Let G be a graph with $n \geq 2$ vertices. As our induction hypothesis, let us assume that the algorithm works for all graphs with $n - 1$ vertices. Let v be the vertex selected in step 2 of the algorithm. Then by the inductive hypothesis, the algorithm outputs $V_1, \ldots, V_k$, an acyclic k partition of $G - v$, where

$k \leq \rho(G - v) \leq \rho(G)$. Let $k' = 1 + \Delta(G) = 1 + \delta(v)$. Let V_i in $V_1, \ldots, V_k$ be such, that v either has no outwards edge to the vertices in V_i, or has no inwards edge from the vertices in V_i. Then, the graph induced by $V_i \cup \{v\}$ is acyclic and we can assign v to V_i to get the valid acyclic partition. If $k \geq k'$ (and hence $\delta(v) < k$), there must exist one such V_i. If there exists no such V_i then it must be the case that $k < k'$ and hence $k < \rho(G)$. We can then create a new set in the partition consisting of the single vertex v. This will be a valid acyclic $\rho(G)$ partition of G.

Finally, it possible to implement the above algorithm in $O(|V| + |E|)$ time, that is, the algorithm can be implemented in linear time. The implementation in Chen (2000) works for the above algorithm *mutatis mutandis* and hence we will not explicitly discuss the linear implementation in this paper. It should be noted however that polynomial time implementations of the above algorithm are fairly straightforward even to the reader not familiar with the advanced data structures used in Chen (2000).

It is worth mentioning that the algorithm outputs the actual partition of vertices not merely the number of elements in the partition. This allows the construction of candidate utility functions for each household member which rationalize the data using the standard techniques in Varian (1982). It is in this way that the algorithm can also be construed to be constructive.

4.2. *Simulations*

We implement the algorithm in MATLAB on a randomly generated data set. We consider a simple situation dependent dictatorship model consisting of a household with 2 decision makers labeled h, w. There are 2 goods priced p_1, p_2 and the quantities of the goods consumed are given by x_1, x_2. Each household member $k \in \{h, w\}$ has the following Cobb Douglas utility function

$$U^k(x_1, x_2) = x_1^{\alpha^k} x_2^{1-\alpha^k}$$

We randomly select α^h, α^w independently from the uniform $\mathbb{U}[0, 1]$ distribution. Then we randomly generate a panel data set consisting of a 100 observations. This is done in the following way. We pick a household member in each observation as the dictator (decision maker) randomly with a probability of $1/2$. We generate prices p_1, p_2 and income I randomly and independently from the uniform $\mathbb{U}[0, 1]$ distribution. Then the consumption bundle x is simply given by $x_1 = \frac{\alpha^k I}{p_1}$

and $x_2 = \frac{(1-\alpha^k)I}{p_2}$ where $k \in \{h, w\}$ is the household member chosen as the dictator for this observation. Notice, this bundle is simply the utility maximizing bundle for household member k. We repeat this process 100 times to generate the data set.

Simulations allow us to test the computational efficiency of the algorithm as well as the tightness of the bound. By randomly generating data, we can construct extremely long panels with substantial variations in prices and incomes which allows us to test the robustness of the algorithm. Simulations suggest that the algorithm is a better approximation on randomly generated data sets with fewer observations per household and on data sets with moderate price variation. Since in practice, it is unlikely to encounter a data set with more than 100 observations per household, the results reported in this section can be viewed as a "lower bound" on the accuracy of the algorithm as it will perform better on data sets with fewer observations. Moreover, real data sets do not contain such large variations in prices and incomes (i.e., have fewer budget set intersections) and hence, potentially have fewer violations of GARP. Also, we allow household members to have vastly divergent preferences. This means that more often than not, the graphs induced by real data will have lower arboricity than our simulated data sets and our algorithm will probably perform even better on them. Lastly, because we construct the data using 2 decision makers we know ex ante that the data set can be partitioned into 2 acyclic components. This provides a benchmark for us to test the performance of the algorithm.

We generated 1000 panel data sets consisting of a 100 observations each. We chose different α^h, α^w for each panel. It took a standard 1.8 GHz Core 2 Duo desktop computer less than three minutes to partition all 1000 data sets. Of the 1000 panels, 175 were partitioned into 1 acyclic component, 662 were partitioned into 2 acyclic components, 159 were partitioned into 3 acyclic components and the remaining 4 were partitioned into 4 acyclic components. Hence, the approximation algorithm gave the correct result 83.7% of the time. Moreover, barring 4 data sets (where it partitioned the graph into 4 components), whenever the algorithm did not give the correct result it gave the next best solution of 3.

5. Conclusion

This paper showed that any arbitrary directed graph can be generated by the revealed preference relation on a data set with the same number of observations as the vertices of the graph in polynomial time. This implies that the revealed preference relation imposes no restriction on the underlying directed graph it generates. This in turn makes the test of the situation dependent dictatorship model NP-Complete. In order to address this issue, we devise a computationally efficent algorithm for computing approximate solutions to the acyclic partitioning problem on directed graphs.

References

[1] Ainslie, A. and Rossi, P. (1998), Similarities in Choice Behavior across Product Categories. *Marketing Science*, 17(2), 91-106.

[2] Bucklin, R. and Gupta, S. (1999), Commercial Use of UPC Scanner Data: Industry and Academic Perspectives. *Marketing Science*, 18(3), 247-273.

[3] Chartrand, G and Kronk, H. (1969), The Point-Arboricity of Planar Graphs. *Journal of the London Mathematical Society*, 44, 612-616.

[4] Chen, Z. (2000), Efficient Algorithms for Acyclic Colorings of Graphs. *Theoretical Computer Science*, 230, 75-95.

[5] herchye, L., De Muynck, T. and De Rock (2010a), Testable Implications of General Equilibrium Models: An Integer Programming Approach. *Working Paper*.

[6] Cherchye, L., De Rock, B. and Vermuelen, F. (2007), The Collective Model of Household Consumption: A Nonparametric Characterization. *Econometrica*, 75(2) , 553-574.

[7] Cherchye, L., De Rock, B., Sabbe, J. and Vermuelen, F. (2008), Nonparametric Tests of Collectively Rational Consumption Behavior: an Integer Programming Procedure. *Journal of Econometrics*, 147, 258-265.

[8] Cherchye, L., De Rock, B. and Vermuelen, F. (2009), The Revealed Preference Approach to Collective Consumption Behavior: Nonparametric Testing and Sharing Rule Recovery. *Review of Economic Studies*, forthcoming.

[9] Cherchye, L., De Rock, B. and Vermuelen, F. (2010b), An Afriat Theorem for the Collective Model of Household Consumption. *Journal of Economic Theory*, 145, 11421163.

[10] Chiappori, P. A. (1988), Rational Household Labor Supply. *Econometrica*, 56(1), 63-90.

[11] Chiappori, P. A.(1992), Collective Labor Supply and Welfare. *Journal of Political Economy*, 100(3), 437-467.

[12] Deb, R. (2010), The Acyclic Partitioning Problem is NP-complete for k = 2.

[13] Dercon, S. and Krishnan, P. (2000), In Sickness and in Health: Risk Sharing within Households in Rural Ethiopia. *Journal of Political Economy*, 108 (4), 688-727.

[14] Dobell, A. (1965), A Comment on A. Y. C. Koo's "An Empirical Test of Revealed Preference Theory". *Econometrica*, 33(2), 451-455.

[15] Duflo, E. and Udry, C. (2004), Intrahousehold Resource Allocation in Cote d'Ivoire: Social Norms, Separate Accounts and Consumption Choices *NBER Working Papers* 10498, National Bureau of Economic Research, Inc.

[16] Famulari, M. (1995), A Household-Based, Nonparametric Test of Demand Theory. *The Review of Economics and Statistics*, 77(2), 372-382.

[17] Garey, M. and Johnson, D. (1979), Computers and Intractability - A Guide to the Theory of NP-completeness. *Freeman*.

[18] Swait, J. and Andrews, R. (2003), Enriching Scanner Panel Models with Choice Experiments. *Marketing Science*, 22(4), 442-460.

[19] Talla Nobibon, F., Cherchye, L., De Rock, B., Sabbe, J. and Spieksma, F. (2010), Heuristics for deciding collectively rational consumption behavior. *Computational Economics*, forthcoming.

[20] Udry, C. (1996), Gender, Agricultural Production, and the Theory of the Household. *Journal of Political Economy*, 104(5), 1010-1046.

[21] Varian, H. (1982), The Nonparametric Approach to Demand Analysis. *Econometrica*, 50(4), 945-974.

[22] Wu, Y., Yuan, J. and Zhao, Y. (1996), Partition a Graph into Two Induced Forests. *Journal of Mathematical Study*, 29(1), 1-6.