Exact inference from finite market data [1]

F. Kübler [2] R. Malhotra [3] H. Polemarchakis [4]

August 22, 2022

[1]This work was circulated earlier as "Identification of preferences, demand and equilibrium with finite data," Discussion paper No. 60, CRETA, Department of Economics, University of Warwick. We thank the referees for many constructive comments.

[2]University of Zürich: fkubler@gmail.com
[3]University of Warwick: r.malhotra@warwick.ac.uk
[4]University of Warwick: h.polemarchakis@warwick.ac.uk

Abstract

We derive conditions under which individual choices can be inferred from finite market data. First, we consider market data that consist of individual demands as prices and incomes change. We show that for any two consumption bundles, strict preference is revealed by finitely many observations of demand. Second, we show that finitely many observations of individual endowments and associated Walrasian equilibrium prices, and only prices, reveal all individuals' strict preferences between two consumption bundles. We explore the implications for forecasting both individual demand and equilibrium prices.

Key words: identification, finite data, preferences, demand, Walrasian equilibrium.

JEL classification: D80; G10.

From only a finite number of observations of market data, can we infer how the individual shall decide when faced with choices not previously encountered? Can we infer equilibrium prices and allocations as individual endowments vary? Do market data need to consist of individual demands, or do observations of aggregate demand or the equilibrium price correspondence suffice for this inference?

Variations of these questions have been extensively studied, but most of the existing literature either focuses on the case of infinitely many observations or poses only the question of whether observations are consistent with utility maximization. The seemingly very important question of what one can conclude about an individual's preferences from finitely many observations has largely been overlooked. In this paper, we pose exactly this question: what does one need to know about preferences a priori in order to be able to make non-trivial, exact inference about the underlying data generating preference with only finitely many data points? Can one predict how an individual will choose from a choice set not previously encountered or predict equilibrium prices at a profile of individual endowments not previously considered?

Revealed preference analysis, the weak axiom, was introduced by Samuelson (1938) as a necessary condition for demand data, a finite collection of pairs of prices and consumption bundles to be generated by the maximization of a preference relation subject to the budget constraint. We trace the development of the literature that followed as a most effective method for understanding where research on the problems stands right now and to highlight the contribution of our paper. Houthakker (1950) introduced the strong axiom, sufficient for demand data to be generated by preference optimization. Later, Afriat (1967) established the generalized axiom of revealed preference as necessary and sufficient for a finite set of demand data to be derived from the maximization of a preference relation or ordinal utility function. Reny (2015) and Nishimura, Ok, and Quah (2017) extended the analysis to arbitrary data sets.

Hurwicz and Uzawa (1971) and Mas-Colell (1977) gave necessary and sufficient conditions for the integrability of a demand function, the derivation of the generating ordinal utility function, and for a demand function to identify preferences. It is worth noting that Lipschitz continuity of preferences or, alternatively, Lipschitz continuity of income expansion paths plays an essential role in the argument.

Brown and Matzkin (1996) extended Afriat (1967) to a finite number of observations on profiles of equilibrium endowments and Walrasian equilibrium prices – that is, observations on the equilibrium manifold. Chiappori, Ekeland, Kubler, and Polemarchakis (2004) showed that the equilibrium manifold locally identifies individual preferences. Brown and Matzkin (1990), and in particular, Matzkin (2006) proved a global version of the result.

While these results on identification answer our motivating question affirmatively for an infinite number of observations, they say nothing for the case of a finite number of observations. They do not even address the question of asymptotics. Mas-Colell (1978) solved this question by considering a nested, increasing sequence of demand data that, at the limit, covers a dense set of prices and provided sufficient conditions for any associated sequence of preference relations to converge to the unique preference relation that generated demand. In a recent paper, Chambers, Echenique, and Lambert (2021) considered the case where data sets are collections of choices from pairwise comparisons of alternatives. Convergence obtains only if the data satisfies a condition implied by, but weaker than monotonicity. For the case of a monotone preference relation, convergence follows from Forges and Minelli (2009).

For a given, finite number of observations, Varian (1982) discussed the possibility of making statements about individual preferences, but he only gave a very partial answer to the problem. In the same framework, Chambers and Echenique (2022) characterized allocations that cannot be rejected as Pareto optimal.

In this paper, we pose the question of inference from a finite number of observations in the following sense: potential observations are infinite, even dense in an appropriate domain, but we require inference to be attained after a finite number of observations. Importantly, the finite number of observations is not a priori fixed. It is this extension of the notion of finite inference that allows for results unattainable with an a priori fixed, even if large, number of observations. And we consider two classes of data: First, we examine the case of observations on an individual's demand at different prices and incomes. Second, for the case of an exchange economy, we assume that only equilibrium prices and profiles of individual endowments are observable[1]. We consider both the *local* cases where observations lie in some open set of

[1]While, in some settings individual consumption is observable, there are other settings where only partial or no observations are available. We derive our results for the extreme case where there are no observations on individual demand.

prices and incomes (in the case of individual demand) or some open set of individual endowments, as well as the *global* case where we allow for all positive prices and incomes, or, in the case of equilibrium, all aggregate endowments and any compact set of income distributions.

In both settings, we consider sequences of nested observations that become dense in the chosen domains as in Mas-Colell (1978)[2]. For both setups, we show that, given any two consumption bundles between which an individual is not indifferent, after sufficiently many observations, we can infer the individual's choices. For the case of individual demand, this fact follows from a result in Mas-Colell (1977), who showed that, if a consumption bundle x is strictly preferred to a bundle y, there must exist prices and incomes such that x is revealed preferred to y. How these revealing prices and incomes are to be found is not addressed. We show that, if x is strictly preferred to y, and if prices and incomes are given, arbitrary, sequences of nested observations that become dense in the chosen domains, then, after finitely many draws, we must observe prices and incomes so that x is strictly revealed preferred to y. From this result, we deduce that for any price vector p, and any $\epsilon > 0$, there are finitely many observations that allow us to determine individual demand at p within ϵ. Our main result is that the situation is identical to the case where only equilibrium prices are observable. If x is strictly preferred to y by some individual, then sufficiently many observations of equilibrium prices and endowment profiles suffice to prove that the individual strictly prefers x to y.

A question that arises directly from this analysis is how these results extend to equilibrium comparative statics. The transfer paradox, introduced by Leontief (1936), makes it clear that knowledge of utility functions is necessary to identify even the direction of welfare effects of transfers. We show that, while predictions of exact comparative statics are generally impossible, we can predict approximate equilibrium prices from finite data. This is true for both setups, independently of whether observations consist of individual choices as prices and incomes vary or of equilibrium prices as the profile of individual endowments varies.

Another obvious question is about possible empirical applications of our results. Blundell (2005) pointed to the advantages of the empirical appli-

[2]Alternatively, observations are drawn randomly from a uniform distribution of exogenous variables, and the endogenous data is generated by utility-maximizing individuals. In this case, all our statements only hold with probability one.

cation of revealed preference theory for measuring consumers' responses to variations in prices and income. Besides the obvious measurement error problem, applied work also needs to address the problem that a failure of the strong axiom does not help characterize the nature of irrationality or the degree/direction of changing tastes – Blundell, Browing, and Crawford (2003) and Blundell, Browing, and Crawford (2008). None of this work considered observations restricted to aggregate demand or equilibrium prices, and it is not clear how these problems can be addressed in that setting.

The results in this paper are therefore not directly applicable in empirical work. However, it is important to identify circumstances where identification is possible in principle. In the absence of identification, empirical non-parametric work is futile.

Our results are comparable to learnability results from statistical learning theory. As Vapnik (2000) put it, *"Between 1960 and 1980 a revolution in statistics occurred; Fisher's paradigm, introduced in the 1920s and 1930s was replaced by a new one. This paradigm reflects a new answer to the fundamental question: What must one know a prior about an unknown functional dependency in order to estimate it on the basis of observations?"* In this vein, Beigman and Vohra (2006) posed the question of whether a demand function derived from a monotone, convex preference relation is *probably, approximately correctly* (PAC) learnable following Valiant (1984). They showed that, without further restrictions on preferences, the set of all demand functions is too rich to be PAC-learnable, while, under the further assumption of Lipschitz continuity, as in Mas-Colell (1977), demand is, indeed, PAC-learnable.

The rest of the paper is organized as follows: Section 1 lays out some preliminaries. Section 2 considers the case where individual demand is observable, and Section 3 focuses on observations of the equilibrium correspondence. An appendix contains the proofs.

1 Preliminaries

We collect definitions and results needed for our analysis.

1.1 Preferences

An individual has a (complete and transitive) preference relation $\succeq$ over the consumption set $\mathbb{R}^L_+$. The preference relation $\succeq$ is upper semi-continuous if, for every $x \in \mathbb{R}^L_+$, the upper contour set $\mathbf{R}_+(x) = \{y : y \succeq x\}$ is closed. It is continuous if, for every $x \in \mathbb{R}^L_+$, the upper contour set as well as the lower contour set $\mathbf{R}_-(x) = \{y : x \succeq y\}$ are closed. It is monotonically increasing (or simply monotone) if $x > y$ implies $x \succ y$. It is convex if $x \succeq y$ implies that $\lambda x + (1-\lambda)y \succeq y$ for all $\lambda \in [0,1]$, and it is strictly convex if $x \succeq y$, $x \neq y$, implies that $\lambda x + (1-\lambda)y \succ y$ for all $\lambda \in (0,1)$.

If, for an open set $\mathbf{X} \subset \mathbb{R}^L_+, x \succeq y \Leftrightarrow x \succeq' y$, we write $\succ =_{\mathbf{X}} \succeq'$.

In the arguments that follow, we assume preferences are differentiably strictly monotonic and strictly convex in the sense of Debreu (1972): They can be represented by a utility function $u : \mathbb{R}^L_+ \to \mathbb{R}$ that is twice continuously differentiable in $\mathbb{R}^L_{++}$, the interior of the consumption set, with $Du(x) \gg 0$ and $D^2u(x)$ negative definite on the orthogonal complement of the gradient $[Du(x)]^\perp$. We refer to such preferences simply as "smooth preferences". This assumption of smooth preferences greatly simplifies many of the arguments, but it should be emphasized that most of our results hold under strictly weaker assumptions. Preferences can be identified from demand functions globally if they are increasing, convex and Lipschitzian following Mas-Colell (1977). Smooth preferences are Lipschitzian, but the converse does not hold. Under the assumption of Lipschitzian preferences, inverse demand functions may not exist, and it is not clear how to define appropriate domains for local identification. Local recoverability was shown under the assumption that preferences generate income Lipschitzian demand in Hurwicz and Uzawa (1971). All smooth preferences satisfy this condition and all preferences that satisfy this condition are Lipschitzian in the sense of Mas-Colell (1977), but there is no equivalence. For our purposes, income Lipschitzian demand functions are not sufficient because our local analysis relies on the existence of an inverse demand function. Diasakos and Gerasimou (2019) proved that inverse demand exists under strictly weaker assumptions than smoothness, but their conditions on preferences do not imply the Lipschitzian property and are not known to guarantee the identification of demand.

1.2 Demand

There are L commodities, $l \in \mathbf{L}$. Budget sets are

$$\mathbf{B}(p) = \{x \in \mathbb{R}^L_+ : p \cdot x \leq 1\}, \quad p \gg 0,$$

and the Walrasian demand correspondence is defined by

$$f^{\succeq}(p) = \{x \in \mathbf{B}(p) : x \succeq y \quad \text{for all } y \in \mathbf{B}(p)\}.$$

For a monotone, strictly convex and continuous preference relation $\succeq$, we use the same notation, $f^{\succeq}(\cdot)$ to denote the (continuous) Walrasian demand function that is generated by $\succeq$. When we want to vary income explicitly, we write

$$d^{\succeq}(p,w) = f^{\succeq}(\frac{p}{w}).$$

Under our assumption of smooth preferences, demand in the interior of the consumption set is characterized by

$$D_x u(x) - \lambda p = 0, \quad p \cdot x = 1, \quad \lambda > 0, \tag{1}$$

and the following properties of demand are easy to derive:

- For $x \in \mathbb{R}^L_{++}$, there is a unique $p \in \mathbb{R}^L_{++}$, such that $x = f^{\succeq}(p)$. That is, there exists an inverse demand function $f^{\succeq -1} : \mathbb{R}^L_{++} \to \mathbb{R}^L_{++}$. From (1), it follows that this function is continuously differentiable.

- For $p = f^{\succeq -1}(x)$, for $x \in \mathbb{R}^L_{++}$, the demand function $f^{\succeq}(p)$ is continuously differentiable and satisfies the Slutsky decomposition

$$D_p d^{\succeq}(p,w) = S^{\succeq}(p,w) - v^{\succeq}(p,w) d^{\succeq}(p,w)', \tag{2}$$

  where the matrix $S^{\succeq}(p,w)$ is symmetric, of rank $(L-1)$, negative semi-definite and satisfies $pS^{\succeq}(p,w) = 0$, and the vector $v^{\succeq}(p,w) = D_w d^{\succeq}(p,w)$ satisfies $pv^{\succeq}(p,w) = 0$.

- For any open, convex set $\mathbf{X} \subset \mathbb{R}^L_+$

$$\succeq =_{\mathbf{X}} \succeq' \quad \Leftrightarrow \quad f^{\succeq -1}(x) = f^{\succeq' -1}(x) \quad \text{for all } \ x \in \mathbf{X}.$$

  As shall become clear below, this local identification result is a necessary condition for most of our results on inference from finite data.

For open sets $\mathbf{P} \subset \mathbb{R}^L_{++}$ and $\mathbf{W} \subset \mathbb{R}_{++}$ we say that a function $f : \mathbf{P} \times \mathbf{W} \to \mathbb{R}^L_+$ is a demand function if it is homogenous of degree zero and if it satisfies Walras law. We say that it satisfies the Slutsky equation if it is continuously differentiable and (2) holds.

1.3 Equilibrium

We consider a pure exchange economy with H individuals, $h \in \mathbf{H}$. We assume consumption sets are $\mathbb{R}^L_+$, and each individual, h has smooth preferences $\succeq^h$, and endowment $e^h \in \mathbb{R}^L_{++}$. A profile of preferences across individuals is $\succeq^{\mathbf{H}}$, and a profile of endowments across individuals is $e^{\mathbf{H}} \in \mathbb{R}^{HL}_{++}$ and aggregate endowments are $e = \sum_{h \in \mathbf{H}} e^h$.

The equilibrium correspondence is defined by

$$\mathbf{\Pi}^{\succeq^{\mathbf{H}}}(e^{\mathbf{H}}) = \{p \in \mathbb{R}^L_{++} : \sum_{h \in \mathbf{H}} (f^{\succeq^h}(\frac{p}{p \cdot e^h}) - e^h) = 0\}.$$

Chiappori, Ekeland, Kubler, and Polemarchakis (2004) and Matzkin (2006) derived sufficient conditions on preferences that ensure classical identification: for any two profiles of preferences $\succeq^{\mathbf{H}}$ and $\tilde{\succeq}^{\mathbf{H}}$,

$$\succeq^{\mathbf{H}} = \tilde{\succeq}^{\mathbf{H}} \Leftrightarrow \mathbf{\Pi}^{\succeq^{\mathbf{H}}}(\cdot) = \mathbf{\Pi}^{\tilde{\succeq}^{\mathbf{H}}}(\cdot).$$

As Balasko (2004) pointed out, without restrictions on the domain of the equilibrium correspondence, this problem becomes trivial since it reduces to the individual problem when individual endowments tend to vanish for all but one individual.

More interestingly, Chiappori, Ekeland, Kubler, and Polemarchakis (2004) and Matzkin (2006) provide a local version of the result. For our setting of finitely many observations, to state the local result, it is useful to observe that this problem is identical to the problem of identification of preferences from aggregate demand.

For individual incomes $w^{\mathbf{H}} = (w^1, \ldots, w^H) \in \mathbb{R}^H_{++}$, we define the aggregate demand function

$$d^{\succeq^{\mathbf{H}}}(p, w^{\mathbf{H}}) = \sum_{h \in \mathbf{H}} f^{\succeq^h}(\frac{p}{w^h}).$$

Note that, for any $e \in \mathbb{R}^L_{++}$ and any profile of incomes $w^{\mathbf{H}} = (w^1, \ldots, w^H)$,

$$p \in \mathbf{\Pi}^{\succeq^{\mathbf{H}}}(\frac{w^1}{\sum_{h \in \mathbf{H}} w^h} e, \ldots, \frac{w^H}{\sum_{h \in \mathbf{H}} w^h} e) \Leftrightarrow d^{\succeq^{\mathbf{H}}}(p, w^{\mathbf{H}}) = e.$$

Therefore, for two profiles of preference relations $\succeq^{\mathbf{H}}$ and $\tilde{\succeq}^{\mathbf{H}}$,

$$d^{\succeq^{\mathbf{H}}}(\cdot) = d^{\tilde{\succeq}^{\mathbf{H}}}(\cdot) \Leftrightarrow \mathbf{\Pi}^{\succeq^{\mathbf{H}}}(\cdot) = \mathbf{\Pi}^{\tilde{\succeq}^{\mathbf{H}}}(\cdot).$$

To state the local version of the result, we define $\mathbf{W}^{\mathbf{H}}$ to be $\times_{h\in\mathbf{H}}\mathbf{W}^h$, where each $\mathbf{W}^h \subset \mathbb{R}_{++}$ is an open set of incomes, and $\mathbf{P} \subset \mathbb{R}^L_{++}$ is an open set of prices: prices and income are observed locally. As Matzkin (2006) pointed out, this differentiates crucially the setting from Balasko (2004), whose argument relied on allowing incomes to approach the boundary. We make the following high-level assumption on preferences and observations introduced in Matzkin (2006).

Assumption 1. For all individuals $h \in \mathbf{H}$ and any preference relation $\succeq' \neq \succeq^h$, that generates a demand function $d^{\succeq'}(.)$ on $\mathbf{P} \times \mathbf{W}^h$ that satisfies the Slutsky equation, there exist $p \in \mathbf{P}$ and $w, w' \in \mathbf{W}^h$, such that

$$d^{\succeq^h}(p, w) - d^{\succeq^h}(p, w') \neq d^{\succeq'}(p, w) - d^{\succeq'}(p, w'). \tag{3}$$

Note that this is a joint assumption on an individual's preferences and on the domain of the demand function, $\mathbf{P} \times \mathbf{W}^h$.

Also, note that the assumption implies directly that for all $\succeq^{\mathbf{H}} \neq \succeq'^{\mathbf{H}}$, there exists a $\bar{p} \in \mathbf{P}$ as well as incomes $\bar{w}^{\mathbf{H}} \in \mathbf{W}$, such that

$$d^{\succeq^{\mathbf{H}}}(\bar{p}, \bar{w}^{\mathbf{H}}) \neq d^{\succeq'^{\mathbf{H}}}(\bar{p}, \bar{w}^{\mathbf{H}}).$$

By the continuity of the aggregate demand function, the assumption implies that, for different profiles of preferences, there must exist an open neighborhood of prices and incomes where aggregate demand (as a function of prices and profiles of income) differs.

The condition in the assumption can only be violated if there is a function $g : \mathbf{P} \to \mathbb{R}^L$, such that, for all $p \in \mathbf{P}$ and $w \in \mathbf{W}$,

$$d^{\succeq}(p, w) = d^{\succeq'}(p, w) + g(p), \quad \text{while} \quad p \cdot g(p) = 0, \tag{4}$$

where the function $g(\cdot)$ is homogeneous of degree zero. If the domain of the demand function satisfies the property that

$$(p, w) \in \mathbf{P} \times \mathbf{W} \Rightarrow \text{for all } \lambda \in (0, 1] : \lambda w \in \mathbf{W}, \tag{5}$$

then Equation (4) can only hold with $g(p) = 0$, and hence Equation (3) must hold automatically. Therefore, it is clear that Assumption 1 must always hold, independently of preferences, when for all individuals h Property (5) is satisfied. This restriction on the domain of the equilibrium correspondence was used in Balasko (2004) to show identification.

The issue at hand is local identification that may fail as the following example shows:

Example 1. Given any utility function $u : \mathbb{R}^L_+ \to \mathbb{R}$ and associated demand function $d^u(p, w)$, define, for $\epsilon \in \mathbb{R}^L$ and $x + \epsilon > 0$,

$$u^\epsilon(x) = u(x_1 + \epsilon_1, \ldots, x_L + \epsilon_L).$$

For $w + \sum_{l=1}^{L} \epsilon_l p_l > 0$, the associated demand function is

$$d^{u^\epsilon}(p, w) = d^u(p, w + p \cdot \epsilon) - \epsilon.$$

If $u(\cdot)$ represents homothetic preferences, $D_w d^{\succeq}(p, w)$ is constant in w, and it is obvious that condition (3) fails. In fact, it is easy to see that Assumption 1 fails for the preference profiles $(\succeq^1, \succeq^2)$ and $(\succeq^{1'}, \succeq^{2'})$, where $\succeq^1$ is identical to $\succeq^2$ and represented by $u(\cdot)$ while $\succeq^{1'}$ is represented by $u^\epsilon(\cdot)$ and $\succeq^{2'}$ is represented by $u^{-\epsilon}(\cdot)$, for some $\epsilon \neq 0$, with $w + p\epsilon > 0$, $w - p\epsilon > 0$ for all $p \in \mathbf{P}$ and all $w \in \mathbf{W}$. In this case, preferences cannot be identified from aggregate demand.

□

By the Slutsky equation and (4) the condition is violated only if there are $\succeq$ and $\succeq'$ with

$$S^{\succeq}(p, w) = D_p d^{\succeq}(p, w) + v(p, w) d^{\succeq}(p, w)',$$

and

$$S^{\succeq'}(p, w) = D_p(d^{\succeq}(p, w) + g(p)) + v(p, w)(d^{\succeq}(p, w) + g(p))'.$$

Since $S^{\succeq}(p, w) - S^{\succeq'}(p, w)$ is a symmetric matrix, the matrix $(v(p, w) - v(p, w'))g(p)'$ must be symmetric for any fixed $p \in \mathbf{P}$ and any $w, w' \in \mathbf{W}$. This implies that, for all w, w', the vector $v(p, w) - v(p, w')$ must be collinear with $g(p)$.

The following proposition summarizes the discussion.

Proposition 1. *If preferences are smooth, Assumption 1 holds if for each $h \in \mathbf{H}$ any of the following conditions hold.*

1. *incomes are unbounded below,*

$$(p, w) \in \mathbf{P} \times \mathbf{W}^h \Rightarrow \textit{for all } \lambda \in (0, 1] : \lambda w \in \mathbf{W}^h,$$

2. *there are $p \in \mathbf{P}$, $w, w', w'' \in \mathbf{W}^h$, $w \neq w'$ and $w \neq w''$, such that, for all $\lambda \neq 0$,*

$$(D_w d^{\succeq^h}(p, w) - D_w d^{\succeq^h}(p, w')) \neq \lambda (D_w d^{\succeq^h}(p, w) - D_w d^{\succeq^h}(p, w'')). \quad (6)$$

Clearly, when preferences are homothetic, condition (6) never holds. Example 1 shows that preferences cannot be identified with local observations. The condition goes a bit further than ruling out just homothetic preferences, but it is not clear how to obtain a clean assumption on preferences that implies (6). It is easy to see that Property (6), together with the assumption that the income effect $v_l^h(\cdot)$ is a twice differentiable function of income w, implies the following condition from Chiappori, Ekeland, Kubler, and Polemarchakis (2004):

1. for every commodity,
$$\frac{\partial v_l^h}{\partial w} \neq 0,$$
while

2. there exist commodities, m and n , such that
$$\frac{\partial}{\partial w}(\ln \frac{\partial v_m^h}{\partial w}) \neq \frac{\partial}{\partial w}(\ln \frac{\partial v_n^h}{\partial w}).$$

Chiappori, Ekeland, Kubler, and Polemarchakis (2004) explained that the assumption relates to the rank condition in Lewbel (1991) which is commonly assumed in applied work.

2 Individual demand

Throughout this section, we assume that we observe choices of a single individual who has preferences $\succeq$ over $\mathbb{R}^L_+$. We consider an infinite sequence of prices $(p_k : k = 1, \ldots)$. We hold the sequence fixed throughout the argument, and we assume that n observations consist of the first n prices of this sequence together with optimal choices, $(p_k, x_k), k = 1, \ldots, n$, where $x_k = f^{\succeq}(p_k)$, for all $k = 1, \ldots, n$. We assume that $\{x_k, k = 1, \ldots\}$ becomes dense in a convex and open set $\mathbf{X} \subset \mathbb{R}^L_+$. Under the assumption that preferences $\succeq$ are smooth, there is an open set of prices $\mathbf{P} = f^{\succeq -1}(\mathbf{X})$, such that $p_k \in \mathbf{P}$ for all k. We consider the “global ”case of $\mathbf{X} = \mathbb{R}^L_{++}$ as well as the local case where $\mathbf{X}$ is bounded.

The two questions we pose are as follows:

1. Given arbitrary $x, y \in \mathbf{X}$, with $x \not\sim y$, can we determine the individual's preferences over the set $\{x, y\}$ after observing some, finite, number n of market choices?

2. Given arbitrary prices $p \in \mathbf{P}$, can we predict with a given degree of precision the individual's demand at these prices after observing some, finite, number n of market choices?

The answers to both questions turn out to be affirmative. But, it is important to point out, as we did earlier, that, posed slightly differently, the first question can lead to negative conclusions: Fixing the number of observations n, there always exist $\{x, y\}$, $x \succ y$, over which one cannot determine the individual's preference. As we show below, things are different concerning the second question: Given any compact set of prices, if the number of observations, n, is large enough, one can approximate the demand function for all prices in the set.

It is well known that a finite number of observations can be rationalized by any strictly convex, continuous, and monotone preference relation if and only if they satisfy the Strong Axiom of Revealed Preference. For completeness, it is useful to state the strong axiom.

Definition 1. Observations satisfy the Strong Axiom of Revealed Preference (SARP) if, for every ordered subset $\{i_1, i_2, ..., i_m\} \subset 1, \ldots$ with $x_{i_k} \neq x_{i_j}$, for all k, j, and with

$$\begin{aligned} p_{i_1} \cdot x_{i_2} &\leq p_{i_1} \cdot x_{i_1}, \\ p_{i_2} \cdot x_{i_3} &\leq p_{i_2} \cdot x_{i_2}, \\ &\vdots \end{aligned}$$

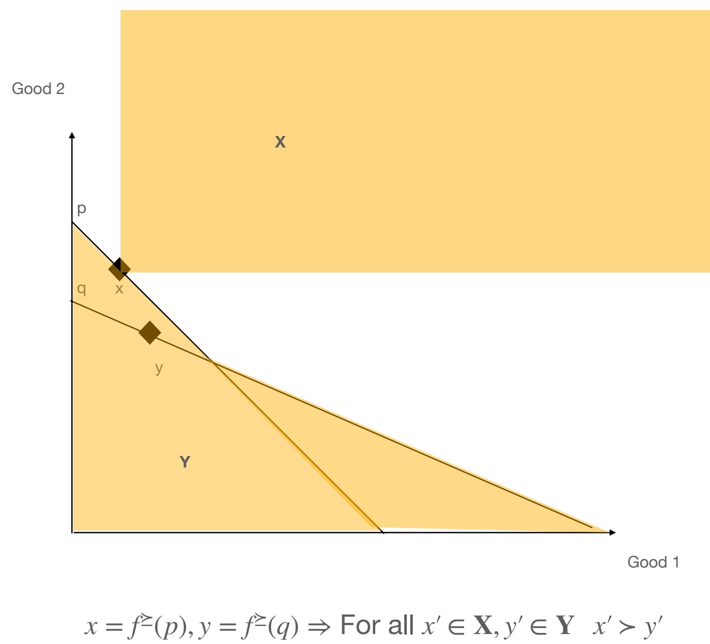
it must be the case that

$$p_{i_m} \cdot x_{i_1} > p_{i_m} \cdot x_{i_m}.$$

Chiappori and Rochet (1987) introduced a strong version of the strong axiom (SSARP) that requires, in addition to SARP, that

$$p_i \neq p_k \Rightarrow x_i \neq x_k, \qquad \text{for all} \ \ i, k \in \{1, \ldots, N\}. \tag{7}$$

They showed that observations can be rationalized by smooth preferences if and only if they satisfy SSARP.



.

Figure 1: From demand to revealed preference

2.1 Identification of pairwise choices

Varian (1982) showed how to construct partial preferences from a finite number of observations on prices and choices. Figure 1 illustrates the basic idea. If we observe that a bundle x is chosen at some prices p and a bundle y is chosen at prices q, and if y lies below the budget line of x, we can infer that x is strictly preferred to y, and, importantly, that all bundles strictly greater than x are strictly preferred to all bundles in the budget set at prices q.

However, the question of whether, given an eventually dense sequence of observations, and any $x \succ y$, there must be some finite number of observations after which x is revealed preferred to y is not addressed in that literature.

For $x, y \in \mathbf{X}$, we say that x is strictly revealed preferred to y through n observations, $x \succ^{R_n} y$, if there are k observations indexed by $i_1, \dots, i_k \in \{1, \dots, n\}$, such that $x \geq x_{i_1}$, $p_{i_j} \cdot x_{i_{j+1}} < p_{i_j} \cdot x_{i_j}$ for all $j = 1, \dots k$ and $x_{i_N} \geq y$. We say x is strictly revealed inferior to y if y is strictly revealed preferred to x. Evidently, if $x \succ^{R_n} y$, for some n, then $x \succ y$. Our first result is a converse: if $x \succ y$, then there is an n such that $x \succ^{R_n} y$.

The result builds on Mas-Colell (1977), Remark 12, and, in parts, our proof closely follows the argument there. The idea of the proof is to show that, for $y \in \mathbf{X}$, the set of $z \in \mathbf{X}$ that are not strictly revealed inferior to y for any n (call this set T_y) is identical to the upper contour set of y – the set of commodity bundles that are weakly preferred to y. In order to do so, we follow Mas-Colell (1977) and construct a new preferences relation $\succeq'$ that is identical to $\succeq$ precisely when T_y is identical to the upper contour set at y. The key is to show that $\succeq'$ generates the same demand function as $\succeq$. Since $\succeq$ is smooth, this implies that the two preference relations must be identical. A formal proof of the following result can be found in the appendix.

Theorem 1. *If $\succeq$ is smooth and $x, y \in \mathbf{X}$, then $x \succ y$ if and only if there is $n \in \{1, \ldots\}$, such that $x \succ^{R_n} y$.*

Mas-Colell (1977) provided, in Example 1, two distinct, strictly convex, monotone, but not smooth, preference relations that generate identical demand functions. In that example identification in our sense is impossible: there exist $x, y \in \mathbf{X}$ and two monotone and strictly convex preference relations $\succeq$ and $\succeq'$, such that

$$f^{\succeq}(p) = f^{\succeq'}(p), \quad \text{for all } p \in \mathbf{P},$$

and $x \succ y$, while $y \succ' x$. Without further assumptions on preferences, identification from market choices is therefore impossible. This seems surprising since, with a finite number of observations, one can always construct smooth preferences if SSARP holds. The theorem shows that the additional condition (7) guarantees that for sufficiently many observations, x is eventually revealed preferred to y if x is strictly preferred to y. As pointed out in Section 1 above, Mas-Colell (1977) shows that the assumption of Lipschitz continuity (that is strictly weaker than smooth preferences) is sufficient for identification. To the best of our knowledge, there is no revealed preference characterization of Lipschitz continuity.

It is instructive to discuss how the theorem relates to the asymptotic results in Mas-Colell (1978). Suppose $x \succ y$ and take any sequence of preferences $\succeq'_k$ that rationalizes $(p_i, x_i)_{i=1}^k$ and satisfies $y \succeq'_k x$. Since we assume that preferences are smooth, it follows from Mas-Colell (1978), Theorem 4 that $\succeq'_k \rightarrow \succeq$ in the topology of closed convergence. When restricted to compact subsets of $\mathbb{R}^L_{++}$ convergence in the topology of closed convergence is equivalent to convergence in the Hausdorff distance (Hildenbrand (1974)).

But, since $x \succ y$, clearly, $(y, x) \in \succeq'_k$ while $(y, x) \notin \succeq$. Furthermore there must be an open ϵ neighborhood around (x, y) in $\mathbf{X} \times \mathbf{X}$ so that $x' \succ y'$ in that neighborhood – a contradiction to the definition of convergence. This convergence argument alone does not directly imply Theorem 1, however. It remains to be shown that there is a non-parametric method to test whether, for a given k, there can be preferences $\succeq'$ that rationalize $(p_i, x_i)_{i=1}^k$ and satisfy $y \succeq' x$. Following Afriat (1967) such a test can be constructed, but this requires an additional argument.

While our result focuses on a single point, it can be easily extended to compact sets. If x is strictly revealed preferred to y there must be an open neighborhood around y such that x is strictly revealed preferred to all points in that neighborhood. If x is strictly preferred to all bundles in some compact set, A, we can find therefore find an open cover and (because of compactness) a finite sub-cover of A consisting of open neighborhoods of points that are strictly revealed inferior to x. This leads to the following result:

Corollary 1. *Suppose preferences are smooth, $A \subset X$ is compact and $x \succ y$ for all $y \in A$, then there is a n, such that $x \succ^{R_n} y$ for all $y \in A$.*

2.2 Identification of demand

For any two bundles between which the individual is not indifferent, a finite number of observations suffice to predict how the individual shall choose. In a market setting, it may be more relevant, however, to ask how the individual shall choose given arbitrary prices and incomes.

Analogously to the analysis above, we can define a revealed demand correspondence as

$$x^{R_n}(p) = \{x \in \mathbf{X} : p \cdot x = 1, \text{ there is no } x' \in \mathbf{X}, p \cdot x' \leq 1, x' \succ^{R_n} x\}.$$

For each p, the set $x^{R_n}(p)$ can be defined by a finite number of linear inequalities that can be computed using Fourier-Motzkin elimination as in Basu, Pollack, and Roy (2006).

It is clear that whenever preferences are convex and monotone, and when $p \in f^{\succeq -1}(\mathbf{X})$,

$$f^{\succeq}(p) \in x^{R_n}(p), \quad n = 1, \ldots.$$

It is also clear that, from n observations on demand, one cannot recover the demand function, and $x^{R_n}(p)$ shall not be single-valued. For sufficiently large n, demand can be arbitrarily well approximated by the revealed demand correspondence: Theorem 1 and strict convexity imply directly that, for any $p \in \mathbf{P}$ and any $x \in \mathbf{B}(p)$ with $x \neq f^{\succeq}(p)$, there must exist $n \in \{1, \ldots\}$, such that $f^{\succeq}(p) \succ^{R_n} x$. Moreover, there must be an open neighborhood around x, such that $f^{\succeq}(p)$ is strictly revealed preferred to all points in the neighborhood. Since the set $\{x \in \mathbf{B}(p) : \|f^{\succeq}(p) - x\| \geq \epsilon\}$ is a compact set, it must have a finite cover of open neighborhoods around points that are strictly revealed inferior to $f^{\succeq}(p)$. This leads to the following result:

Theorem 2. *Suppose preferences are smooth. Given any compact $\widetilde{\mathbf{P}} \subset \mathbf{P}$ and any $\epsilon > 0$, there exists $n \in \{1, \ldots\}$, such that, for all $p \in \widetilde{\mathbf{P}}$,*

$$x^{R_n}(p) \subset \{x \in \mathbf{B}(p) : \|f^{\succeq}(p) - x\| \leq \epsilon\}.$$

Note that, given $n \in \{1, \ldots\}$, observations, the $\epsilon > 0$ in the statement of the result can be computed since the set $x^{R_n}(p)$ can be explicitly computed.

3 Equilibrium

So far, our analysis has focused on the classical demand setting. More generally, economic theory derives relationships between the fundamentals of the economy, some of which may not be observable, and observed individual or aggregate behaviour or equilibrium prices. It is then of interest to ask whether what is observed can be used to deduce individual preferences. We tackle this issue in the classical setting of Walrasian equilibrium and assume that individual behavior, that is, individual demand, is not observable.

Observations consist only of equilibrium prices and individual endowments or equivalently, as we showed in Section 1, of prices, individual incomes, and the resulting aggregate demand. As mentioned in the introduction, one could imagine a setting where individual demand is partially observed. Epstein (1982) examined integrability of individual demand in such a setting (without observing aggregate demand). Clearly partial observations on demand do generally not allow for identification, whether or not they would allow us to relax Assumption 1 is subject to further research.

Throughout the section, we assume that there are H individuals, $h \in \mathbf{H}$ with preferences $\succeq^h$, for each h. Analogously to our analysis above, we are interested in identifying preferences for individual h over a convex and open set $\mathbf{X}^h \subset \mathbb{R}^L_+$. However, now we are in a setting where we do not observe individual choices and hence cannot "observe" $\mathbf{X}^h$. We, therefore, assume throughout this section that aggregate demand (or the equilibrium correspondence) is observed at prices and incomes (or profiles of individual endowments) where every individual's demand is strictly positive[3].

We consider a sequence of prices and individual incomes. We assume that $\{(p_k, w_k^1, \ldots, w_k^H), \quad k = 1, \ldots\}$ become dense in $\mathbf{P} \times \mathbf{W^H} \subset \mathbb{R}_{++}^L \times \mathbb{R}_{++}^H$, and we define observations of aggregate demand as

$$D_k = d^{\succeq^{\mathbf{H}}}(p_k, w_k^1, \ldots, w_k^H), \quad k = 1, \ldots.$$

Given n observations, it is useful to define the set of consumption allocations that are consistent with the n observations in the sense that they add up to aggregate demand and satisfy the strong axiom:

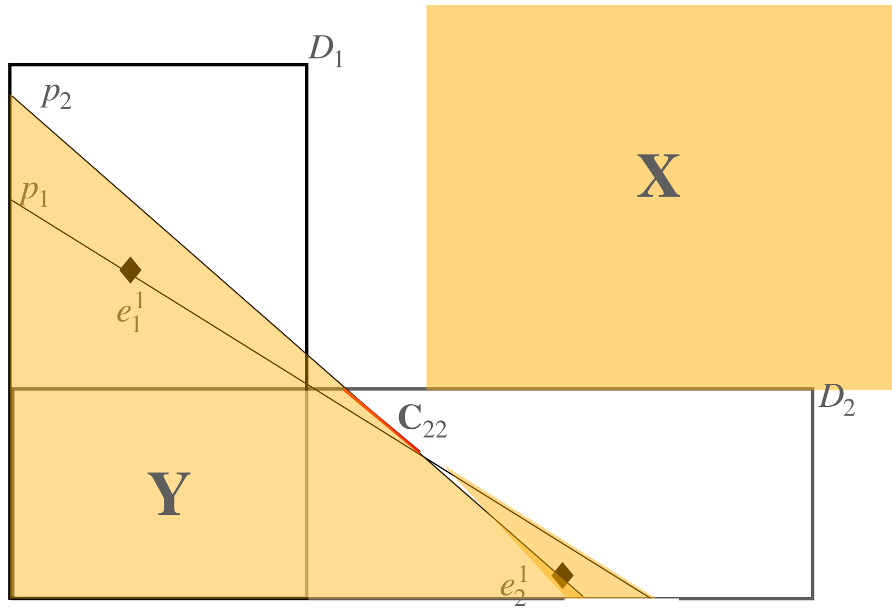
$$\mathbf{C}_n = \left\{ x \in \mathbb{R}_+^{nHL} : \begin{array}{c} \sum_{h \in \mathbf{H}} x_k^h = D_k, \quad k = 1, \ldots, n, \\ \\ p_k \cdot x_k^h = w_k^h, \quad k = 1, \ldots, n, \ \text{for all } h \in \mathbf{H} \\ \\ (x_k^h, p_k)_{k=1}^n \quad \text{satisfy SARP for all } h \in \mathbf{H} \end{array} \right\}.$$

Note that this is a semi-algebraic set that can be written as the finite union and intersection of sets of the form $\{x \in \mathbb{R}^n : g(x) > 0\}$ or $\{x \in \mathbb{R}^n : f(x) = 0\}$, where f and g are polynomials in x with coefficients in $\mathbb{R}$. As in Basu, Pollack, and Roy (2006), quantifier-elimination can be used to compute the sets.

We denote the projection of $\mathbf{C}_n$ onto the coordinates corresponding to the k'th observation by $\mathbf{C}_{nk} \subset \mathbb{R}_+^{HL}$, $k = 1, \ldots, n$. That is, $\mathbf{C}_{nk}$ is the set of all consumption vectors across the H individuals that are on the budget hyperplane and add up to aggregate consumption, and for which there are consumptions for all other observations $k' \neq k$ that also add up to the relevant aggregate consumptions, are on relevant budget planes and, crucially, altogether satisfy the strong axiom of revealed preference.

We illustrate how observations on aggregate demand, individual endowments and equilibrium prices limit the set of possible consumption vectors. Figure 2 shows two Edgeworth-boxes, D_1 corresponding to prices p_1 and endowments $(e_1^1, D_1 - e_1^1)$ and D_2 corresponding to prices p_2 and endowments $(e_2^1, D_2 - e_2^1)$. The red line-segment, C22 on the budget line p_2 corresponds

[3] Debreu (1972)'s assumption that $\mathbf{R}_+(x) \subset \mathbb{R}_{++}^L$, for any $x \in \mathbb{R}_{++}^L$, guarantees that demand is interior at all prices $p \gg 0$.



.

Figure 2: From pairwise choice to revealed preference

to the possible consumption of individual 1 in the second observation. Since individual 1 must consume insight the first Edgeworth box given prices p_1, the only consumption consistent with the weak axiom lies on the red line-segment, C22. We can then conclude that any bundle in the set $\mathbf{X}$ is strictly preferred to any bundle in $\mathbf{Y}$.

Importantly, in our context, Assumption 1 implies that, for sufficiently large n, the sets $\mathbf{C}_{nk}$ become *small* in the sense that they contain only a neighborhood around the profile of actual individual demands. The key to this result lies in the fact that if for some k and some h, some $y_k^h \neq f^{\succeq^h}(\frac{p_k}{w_k^h})$ then there must be an n such that $y_k^{\mathbf{H}} \notin \mathbf{C}_{nk}$. If $y_k^{\mathbf{H}} \in \mathbf{C}_{nk}$ for all n, since SARP holds, there must be some preferences $\succeq^{h\prime}$ that rationalize the data (see Richter (1966)) and this contradicts Assumption 1. An argument similar to the one in the proof of Theorem 2 then leads to the following result:

Lemma 1. *Suppose Assumption 1 holds and preferences are smooth. Given any $\epsilon > 0$, and any observation k, there exists an n such that*

$$x^{\mathbf{H}} \in \mathbf{C}_{nk} \Rightarrow \|x^h - f^{\succeq^h}(\frac{p_k}{w_k^h})\| < \epsilon \text{ for all } h \in \mathbf{H}.$$

As was the case in Theorem 2, for a given n the ϵ can be calculated explicitly since $\mathbf{C}_n$ can be computed explicitly.

The result is surprising: Individual demands can be identified from the aggregate demand function! However, note that the result is consistent with the results in Chiappori, Ekeland, Kubler, and Polemarchakis (2004) and in particular in Matzkin (2006). Observation of aggregate demand means that one observes the income effects of each individual's demand. The contribution of our lemma is to show that with finite data, this is approximately true.

3.1 Identification of individual choice

As in Section 2 we want to develop a notion of "equilibrium revealed preferred", meaning that, through observations of equilibrium prices, it is revealed that an individual must prefer some bundle x to another bundle y. Given $x, y \in \mathbf{X}$, we say that x is equilibrium revealed preferred to y by individual h given n observations if for all $(x_k^{\mathbf{H}})_{k=1,\ldots,n} \in \mathbf{C}_n$, there are N observations indexed by $i_1, \ldots, i_N \in \{1, \ldots, n\}$, so that $x \geq x_{i_1}^h$, and $p_{i_j} \cdot x_{i_{j+1}}^h < p_{i_j} \cdot x_{i_j}^h$, for all $j = 1, \ldots N$, as well as $x_{i_N}^h \geq y$. If x is equilibrium revealed preferred to y by an individual h we write $x \succ^{R_n^h} y$.

While for the classical notion of revealed preferred, x is said to be revealed preferred to y if choices reveal that an individual must prefer x to y, in this setting, equilibrium prices reveal that an individual prefers x to y. This is possible because incomes vary, and Lemma 1 shows that individual choices can be recovered (approximately) from aggregate demand.

Obviously, if there is a n such that $x \succ^{R_n^h} y$ then $x \succ^h y$. As in Section 2, we show that the converse also holds, for sufficiently large n, $x \succ^{R_n^h} y$ must hold whenever $x \succ^h y$. The result follows by combining Theorem 1 with Lemma 1.

Theorem 3. *Suppose individual preferences are smooth and Assumption 1 is satisfied. Given any $h \in \mathbf{H}$, there is an open and convex set $\mathbf{X}^h$ such that, for any $x, y \in \mathbf{X}^h$ with $x \succ^h y$, there exists an n, such that*

$$x \succ^{R_n^h} y.$$

This is the main result of our paper. Sufficiently many observations on the equilibrium manifold allow us to infer how any one of the individuals in the economy will choose between two bundles.

The theorem shows the existence of open and convex sets $\mathbf{X}^h$ for each individual h. When $\mathbf{P} = \mathbb{R}^L_{++}$ the set $\mathbf{X}^h = \mathbb{R}^L_{++}$.

It follows directly from the proof of Theorem 2 above that sufficiently many observations on the equilibrium manifold also allow us to predict Walrasian demand of individual h at prices $p \in f^{\succeq^h -1}(\mathbf{X}^h)$. Formally we can define the equilibrium revealed demand correspondence as

$$x^{R^h_n}(p) = \{x \in \mathbf{X}^h : p \cdot x = 1, \text{ there is no } x' \in \mathbf{X}^h, p \cdot x' \leq 1, x' \succ^{R^h_n} x\}.$$

Note that the definition is identical to the definition in Section 2 except that we replace "revealed preferred" with "equilibrium revealed preferred". and that the domain of the revealed demand correspondence is only implicitly defined. As above, it is clear that whenever preferences are convex and monotone, for each agent h

$$f^{\succeq^h}(p) \in x^{R^h_n}(p) \quad n = 1, \ldots.$$

The following result shows that for sufficiently large n demand can be arbitrarily well approximated by the revealed demand correspondence:

Theorem 4. *Suppose that preferences are smooth and that Assumption 1 holds. Given any individual h and compact $\widetilde{\mathbf{P}} \subset f^{\succeq^h -1}(\mathbf{X}^h)$ and any $\epsilon > 0$ there exists an $n \in \{1, \ldots\}$, such that for all $p \in \widetilde{\mathbf{P}}$,*

$$x^{R^h_n}(p) \subset \{x \in \mathbf{B}(p) : \|f^{\succeq}(p) - x\| \leq \epsilon\}.$$

Given the result in Theorem 2 and the proof of Theorem 3, the result follows immediately.

For the case of $\mathbf{P} \neq \mathbb{R}^L_{++}$ the result is a bit unsatisfactory since the domain where demand can be approximated well is not explicitly given.[4] In the last subsection we therefore focus on the case $\mathbf{P} = \mathbb{R}^L_{++}$.

3.2 Identification of Equilibrium

A more delicate issue is whether one can forecast equilibrium prices. A resolution of the transfer paradox would require this. An obvious problem that arises is that the possibility of multiple equilibria cannot easily be ruled out. But perhaps, even in the presence of multiplicity, one can predict one of the

[4]Note, however, that it can be computed from the observations.

equilibria. Unfortunately, things become much more complicated because of the intricate relation between approximate equilibria and exact equilibrium. While it is true that for any economy and any $\delta > 0$ there is ϵ, such that, if the norm of aggregate excess demand is smaller than ϵ, prices are within δ of equilibrium prices (as in Anderson (1986)), there is no constructive algorithm to determine ϵ. In fact, Richter and Wong (1999) suggested that this is generally the case. We will therefore focus on approximate equilibria.

Given n observations of prices and incomes, $(p_k, w_k^{\mathbf{H}})_{k=1,\dots,n}$, with $p_k \in \mathbb{R}^L_{++}$, and $w_k^{\mathbf{H}} \in \mathbf{W}^{\mathbf{H}}$ for all $k = 1, \dots, n$ we define $\mathbf{P}(e^{\mathbf{H}}) = \{p \in \mathbb{R}^L_{++} : p{\cdot}e^h \in \mathbf{W}^h$ for all $h \in \mathbf{H}\}$ and the revealed approximate equilibrium correspondence as follows:

$$\mathbf{\Pi}^{R_n^{\mathbf{H}}}(e^{\mathbf{H}}) = \left\{ p \in \mathbf{P}(e^{\mathbf{H}}) : \exists\, x^{\mathbf{H}} \in \mathbb{R}^{HL}_{++}, \quad \begin{array}{c} x^h \in x^{R_n^h}(\frac{p}{w_h}) \text{ for all } h \in \mathbf{H}, \\ \sum_{h \in \mathbf{H}} (x^h - e^h) = 0, \end{array} \right\}.$$

As above, it is clear that for a given profile of individual endowments, $e^{\mathbf{H}}$, the set of equilibria with prices in $\mathbf{P}(e^{\mathbf{H}})$ is a subset of $\mathbf{\Pi}^{R_n^{\mathbf{H}}}(e^{\mathbf{H}})$. It follows directly from Theorem 4 that, for any $\epsilon > 0$, the correspondence describe the set of all ϵ-equilibria with incomes in $\mathbf{W}^{\mathbf{H}}$ and with prices in $\mathbf{P}(e^{\mathbf{H}})$. Our final result is as follows:

Theorem 5. *Suppose $\mathbf{P} = \mathbb{R}^L_{++}$, Assumption 1 holds, and preferences are smooth. For all $e^{\mathbf{H}} \in \mathbb{R}^{HL}_{++}$,*

$$\mathbf{\Pi}^{\succeq^{\mathbf{H}}}(e^{\mathbf{H}}) \cap \mathbf{P}(e^{\mathbf{H}}) \subset \mathbf{\Pi}^{R_n^{\mathbf{H}}}(e^{\mathbf{H}}).$$

Furthermore, for any $\epsilon > 0$ and any compact $\mathbf{E} \subset \mathbb{R}^{HL}_{++}$ there exists $n \in \{1, \dots\}$, such that, for all $e^{\mathbf{H}} \in \mathbf{E}$,

$$\mathbf{\Pi}^{R_n^{\mathbf{H}}}(e^{\mathbf{H}}) \subset \{p \in \mathbf{P}(e^{\mathbf{H}}) : \| d^{\succeq^{\mathbf{H}}}(p, p \cdot e^1, \dots, p \cdot e^H) - e \| < \epsilon\}.$$

As explained above, due to the fact that small perturbations of fundamentals can have large effects on equilibrium prices, it is difficult to go beyond this result and make statements about exact equilibria. When it is known a priori that equilibrium is unique for all profiles of endowments, the set of approximate equilibrium prices must shrink to consist, eventually, only of a neighborhood of the exact equilibrium, and sufficiently many observations allow us to identify equilibrium prices within ϵ.

A Appendix: Proofs

Proof of Theorem 1

The "if" part is clear.

For the converse, given any $y \in int(\mathbf{X})$, define

$$T_y = \{z \in \mathbf{X} : \text{there is no } n, \text{such that } y \succ^{R_n} z\},$$

and define a new preference relation $\succeq'$ by

$$u \succeq' v \text{ if } \begin{cases} u \in \text{conv}(T_y \cup \mathbf{R}_+(v)) & \text{if } v \notin T_y \\ \\ u \in T_y \cap \mathbf{R}_+(v) & \text{if } v \in T_y \end{cases},$$

where $\mathbf{R}_+(\cdot)$ denotes the upper contour set under the preference relation $\succeq$.

It can be verified that $\succeq'$ is upper-semi continuous, monotone and convex, and we can define the demand correspondence $f^{\succeq'}$. We shall argue below that it suffices to show that $f^{\succeq}(\cdot) = f^{\succeq'}(\cdot)$. To prove this, note first that $f^{\succeq'}(\cdot)$ is non-empty for all p. Then suppose that for some $p \in \mathbf{P}$, $u \neq v = f^{\succeq}(p)$ but $u \in f^{\succeq'}(p)$. By the definition of $\succeq'$, this can only be the case if $u \in T_y$ but $v \notin T_y$. By the continuity of $f^{\succeq}(\cdot)$, for sufficiently large number of observations, n, there must be a $j \in \{1, \ldots, n\}$ and a p'_j sufficiently close to p so that $f^{\succeq}(p) \notin T_y$ but $p'_j \cdot f^{\succeq}(p'_j) > p'_j \cdot u$. This is a contradiction to transitivity since we would have that y is revealed preferred to $f^{\succeq}(p'_j)$ which is revealed preferred to u and u cannot be in T_y.

Therefore, $\succeq$ and $\succeq'$ generate the same demand functions on $\mathbf{P}$. This implies that the two preference relations coincide, which is only possible if T_y is equal to the upper contour set of $\succeq$ at y.

Proof of Theorem 2

For any $p \in f^{\succeq -1}(\mathbf{X})$, since preferences are strictly convex, $f^{\succeq}(p) \succ x$, for all $x \in \mathbf{B}(p), x \neq f^{\succeq}(p)$. By Theorem 1, there exist some n so that $f^{\succeq}(p) \succ^{R_n} x$. Take any compact set

$$\mathbf{K} \subset \{(p, x) : p \in \mathbf{P}, x \in \mathbf{B}(p) \setminus \{f^{\succeq}(p)\}\}.$$

Whenever $f^{\succeq}(\bar{p}) \succ^{R_n} \bar{x}$, there must be an open set around $(\bar{p}, \bar{x})$ so that for any (p, x) in that set $f(p)$ is revealed preferred to x. Therefore, there is a collection of open sets covering $\mathbf{K}$. Since $\mathbf{K}$ is compact, there exists a finite subcover. Each of the finitely many points defining the subcover is observed at some finite number of observations which proves the result.

Proof of Lemma 1

Suppose that, for some k, $y_k^{\mathbf{H}} \in \mathbf{C}_{nk}$ for all n but that for some h, $y_k^h \neq f^{\succeq^h}(\frac{p_k}{w_k^h})$. By the definition of $\mathbf{C}_{nk}$ there is an infinite sequence of prices and choices (y_l^h, p_l), $l = 1, \ldots$, that satisfy SARP and therefore each $y_l^h = f^{\succeq^h\prime}(\frac{p_l}{w_l^h})$ for some preference relation $\succeq^{h\prime}$ (see Richter (1966)). Under the assumption that all agents' preferences $\succeq^{\mathbf{H}}$ are smooth, individual demand is differentiable and satisfies the Slutksy equation. If $d^{\succeq^{\mathbf{H}\prime}}(.) = d^{\succeq^{\mathbf{H}}}(.)$ on an open set $\mathbf{P} \times \mathbf{W}^{\mathbf{H}}$ then each $d^{\succeq^{h\prime}}(.)$ must be continuously differentiable and since SARP holds the Slutksy equation must hold (see Kihlstrom et al. (1976)). Since $\succeq^{h\prime} \neq \succeq^h$, by Assumption 1, this implies that there must be some observation j for which

$$d^{\succeq^{\mathbf{H}}}(p_j, w_j^{\mathbf{H}}) \neq d^{\succeq^{\mathbf{H}\prime}}(p_j, w_j^{\mathbf{H}}).$$

This contradicts the assumption that $y_k^{\mathbf{H}} \in \mathbf{C}_{nk}$ for all n, and, as a consequence, there must exist a $\overline{n}$ such that $y_k^{\mathbf{H}} \notin \mathbf{C}_{\overline{n},k}$.

For all $n \leq \overline{n}$, if some $(y_k^{\mathbf{H}})_{k=1 \ldots n-1} \in \mathbf{C}_{n-1}$ but there is no $y_n^{\mathbf{H}}$ so that

$$(y_k^{\mathbf{H}})_{k=1 \ldots n} \in \mathbf{C}_n,$$

then for all $y_n^{\mathbf{H}}$ with

$$\sum_{h \in \mathbf{H}} y_n^h = D_n,$$

there must be some h and a budget feasible $\tilde{y}^h$ that is strictly revealed preferred to y^h. Since this is a strict order, this must be true for an open neighborhood of $y_k^{\mathbf{H}}$. Since the intersection of finitely many open neighborhoods forms an open neighborhood, there is an open neighborhood of $y_k^{\mathbf{H}}$ that is not contained in $\mathbf{C}_{\overline{n},k}$.

The same argument (with possibly different $\overline{n}$) applies to all $y_k^{\mathbf{H}}$ with $y_k^h \in \{x \in \mathbf{B}(\frac{p_k}{w_k}) : \|x - f^{\succeq^h}(\frac{p_k}{w_k})\| \geq \epsilon\}$ for some h. Since for each h, $\{x \in \mathbf{B}(\frac{p_k}{w_k}) : \|x - f^{\succeq^h}(\frac{p_k}{w_k})\| \geq \epsilon\}$ is a compact set, the open cover generated by all these $y^{\mathbf{H}}$ has a finite sub-cover and hence there must be some finite n^* that gives the result.

Proof of Theorem 3

For each agent h there is an open and convex $\mathbf{X}^h$ such that $f^{-1}(\mathbf{X}^h) \subset \mathbf{P}$. For any $x, y \in \mathbf{X}^h$, Theorem 1 implies that there is an $\bar{n}$ such that for all $n > \bar{n}$ there is a $(x_k^{\mathbf{H}})_{k=1,\ldots,n} \in \mathbf{C}_n$ so that the associated (x_k^h) to together with (p_k) imply that x is revealed preferred to y by individual h. Lemma 1 implies that for sufficiently large n there can be no other solutions $(\tilde{x}_k^{\mathbf{H}})_{k=1,\ldots,n} \in \mathbf{C}_n$ for which this does not hold.

References

S. Afriat. The construction of a utility function from demand data. *International Economic Review*, 8:67–77, 1967.

R. M. Anderson. "Almost" implies "near". *Transactions of the American Mathematical Society*, 296(1):229–237, 1986.

Y. Balasko. The equilibrium manifold keeps the memory of individual demand functions. *Economic Theory*, 24(3):493–501, 2004.

S. Basu, R. Pollack, and M.-F. Roy. *Algorithms in real algebraic geometry*, volume 10. Springer, 2006.

E. Beigman and R. Vohra. Learning from revealed preference. In *Proceedings of the 7th ACM Conference on Electronic Commerce*, pages 36–42, 2006.

R. Blundell. How revealing is revealed preference? *Journal of the European Economic Association*, 3:211–235, 2005.

R. Blundell, M. Browing, and I. Crawford. Nonparametric engel curves and revealed preference. *Econometrica*, 71:205–240, 2003.

R. Blundell, M. Browing, and I. Crawford. Best nonparametric bounds on demand responses. *Econometrica*, 76:1227–1262, 2008.

D. Brown and R. Matzkin. Recoverability and estimation of the demand and utility functions of traders when demands are unobservable. Unpublished manuscript, 1990.

D. Brown and R. Matzkin. Testable restrictions on the equilibrium manifold. *Econometrica*, 64:1249–1262, 1996.

Ch. Chambers and F. Echenique. Empirical welfare economics. Unnpublished manuscript, 2022.

Ch. Chambers, F. Echenique, and N. Lambert. Recovering preferences from finite data. *Econometrica*, 89(4):1633–1664, 2021.

P.-A. Chiappori and J.-C. Rochet. Revealed preferences and differentiable demand. *Econometrica*, 55:687–691, 1987.

P.-A. Chiappori, I. Ekeland, F. Kubler, and H. Polemarchakis. Testable implications of general equilibrium theory: a differentiable approach. *Journal of Mathematical Economics*, 40:105–119, 2004.

G. Debreu. Smooth preferences. *Econometrica*, 603-615:40, 1972.

Th. Diasakos and G. Gerasimou. Preference conditions for invertible demand functions. Available at SSRN 2921099, 2019.

L.G. Epstein. Integrability of incomplete systems of demand functions. *The Review of Economic Studies*, 49(3):411–425, 1982.

F. Forges and E. Minelli. Afriat's theorem for general budget sets. *Journal of Economic Theory*, 144(1):135–145, 2009.

W. Hildenbrand. *Core and Equilibria of a Large Economy.* Princeton University Press, 1974.

H. S. Houthakker. Revealed preference and the utility function. *Economica*, 17:159–174, 1950.

L. Hurwicz and H. Uzawa. On the integrability of demand functions. In L. Hurwicz J. Chipman and H. Sonnenschein, editors, *Preferences, Utility and Demand*. Harcourt Brace, Jovanovich, 1971.

R. Kihlstrom, A. Mas-Colell, and H. Sonnenschein. The demand theory of the weak axiom of revealed preference. *Econometrica*, 44:971–978, 1976.

V. Leontief. A note on the pure theory of transfer. In *Exploration in Economics: Notes and Essays in Honor of F. W. Taussig*. McGraw Hill, 1936.

A. Lewbel. The rank of demand systems: theory and nonparametric estimation. *Econometrica*, 59:711–730, 1991.

A. Mas-Colell. On the recoverability of consumers' preferences from demand behavior. *Econometrica*, 45:1409–1430, 1977.

A. Mas-Colell. On revealed preference analysis. *The Review of Economic Studies*, 45:121–131, 1978.

R. L. Matzkin. Identification of consumer preferences when their choices are unobservable. In *Rationality and Equilibrium*, pages 195–215. Springer, 2006.

H. Nishimura, E. A. Ok, and J. K.-H. Quah. A comprehensive approach to revealed preference theory. *American Econoimc Review*, 107.:1239–1263, 2017.

Ph. J. Reny. A characterization of rationalizable consumer behavior. *Econometrica*, 83:175–192, 2015.

M. K. Richter. Revealed preference theory. *Econometrica*, 34:635–645, 1966.

M.K. Richter and K.-C. Wong. Non-computability of competitive equilibrium. *Economic Theory*, 14(1):1–27, 1999.

P. A. Samuelson. A note on the pure theory of consumers' behavior. *Economica*, 5:61–71, 1938.

G. Valiant. A theory of the learnable. *Communications of the Association for Computing Machinery*, 27:1134–1142, 1984.

V. N. Vapnik. *The Nature of Statistical Learning Theory.* Springer Verlag, 2nd edition, 2000.

H. R. Varian. The nonparametric approach to demand analysis. *Econometrica*, 50:945–973, 1982.