

Naturally Occurring Preferences and General Equilibrium: A Laboratory Study*

Sean Crockett

Daniel Friedman

Ryan Oprea¹

*Chapman University, USA; University of California at Santa Cruz, USA; University of
California at Santa Barbara, USA*

We examine whether economies constructed using experimentally measured preferences (for risk) tend to suffer from aggregation pathologies (like non-existence and multiplicity of equi-

***Submitted:** June 6, 2020. **Revised:** July 11, 2020; October 30, 2020. **Finalized:** December 10, 2020

¹We are grateful to the National Science Foundation for support under grant SES-1357867, to Emily Hockel and Matt Jee for programming assistance, and to Akhtar Shah and Roman Shelkov for research assistance. We are also grateful to Douglas Gale, Bill Zame, Elena Asparouhova, Peter Boessaerts, Colin Camerer, John Duffy, John Ledyard, Steve Spear, Olivier Armantier, Luke Lindsay, Tibor Neugebauer, Brian Rogers, PJ Healy and Jan Werner for helpful comments, and to seminar audiences at UCLA, Columbia University, Ohio State University, University of Utah Finance, Humboldt University Berlin, University of Melbourne Finance, University of Oregon, the 2017 Experimental Economics session of the Stanford Institute for Theoretical Economics, the 2018 Economic Science Association Meetings in Berlin, the Southwestern Experimental and Behavioral Economies workshop, the 2018 Society for Experimental Finance Meetings in Heidelberg, the 2016 North American meetings of the Economic Science Association, the 2017 Society for the Advancement of Economic Theory meetings in Faro, the 2018 Neuroeconomics and the Evolution of Economic Behavior Workshop in Vancouver, the 2019 Conference on the Experimental and Behavioral Aspects of Financial Markets at Chapman University, the 2019 Financial Theory and Experiments Conference at NYU.

librium) cautioned by the Sonnenschein-Mantel-Debreu theorem. We show that aggregation pathologies should be expected to arise frequently in homogeneous exchange economies, but dwindle and eventually disappear as economies grow diverse. When subjects actually trade, general equilibrium predictions are accurate in most cases, and the failures occur only in economies a priori classified as fragile to preference instability. Our study uses individual-level experimental data to address longstanding questions in general equilibrium theory that are unanswerable using prior methods.

Running Title Header: Preferences and General Equilibrium

Keywords: Experimental Economics, General Equilibrium, Aggregation, Heterogeneity, Revealed Preferences.

JEL codes: C91, C92, D03, D51, G11

1 Introduction

In the 1970s it became clear that general equilibrium theory, at that time widely regarded as the centerpiece of economics, suffers from two fundamental problems. First, as the Sonnenschein-Mantel-Debreu theorem (SMD) (Sonnenschein, 1973; Mantel, 1974; Debreu, 1974) showed, even very standard neoclassical preferences can aggregate to produce “pathological” economies, for which competitive equilibrium is non-existent, or multiple, or dynamically unstable. Second, the theory itself tells us very little about how, when and why markets should converge to competitive equilibria even when they do exist. Taken together, these problems call into question the power of the theory to predict and interpret market behavior.

In this paper we introduce new experimental methods that shed new empirical light on those longstanding problems. Our contribution is to study laboratory economies that are built entirely

from the “homegrown” preferences that subjects bring with them into the lab. The first phase of our experiment has each subject reveal his or her preferences by choosing an allocation from each of dozens of exogenously given budget lines, each generated by a fixed initial endowment and a grid of prices.² This allows us to study the range of pure exchange economies that can arise, given the nature and diversity of preferences we observe among subjects. Each possible grouping of subjects’ preferences defines an *economy* and for each, we can easily and nonparametrically calculate aggregate excess demand at grid prices. Those excess demands directly yield “revealed competitive equilibria” (RCE), and their structure enables us to study for the first time whether aggregations of actual preferences tend to give rise to the sort of predictive pathologies highlighted in SMD.

The second phase of each laboratory session produces actual *market outcomes* for four of the many economies that can be constructed from the set of preferences we observe. Specifically, we use a sorting algorithm to group subjects into four specific economies, and then have subjects use the tatonnement trading institution to produce final allocations and prices, which we compare to the RCE predictions. The sorting algorithm produces economies whose RCE predictions span a wide range of prices and final allocations, enabling demanding tests of the predictive power of general equilibrium theory over market outcomes.³

These methods contrast with the prior experimental general equilibrium literature which, in-

²Using a fixed initial endowment and grid of prices is a slight but (for our purposes) important difference from Choi, Fisman, Gale and Kariv (2007) who vary both endowment and price across budget lines.

³We emphasize the distinction between a general equilibrium “economy” (a description of the aggregate properties of a collection of preferences), and a “market” (actual trading behavior in any such economy). By examining the various ways preferences can combine, we can study the range of economies implicit in a sample of preferences and their properties. By actually having a set of subjects trade under a market institution, we can compare resulting market outcomes to an equilibrium of the economy formed from that set of subjects’ preferences.

stead, generally studies market behavior in *induced preference* economies. In such experiments (discussed in more detail below) subjects are (i) endowed with bundles of fictional goods, (ii) given artificial preferences in the form of non-linear monetary reward functions over their final allocation (i.e. the bundle they hold after trading in a market), and (iii) allowed to trade goods using a market institution. Observed trade dynamics and outcomes then can be compared to the equilibrium predictions for the endowments and induced preferences assigned by the experimenter. These methods have helped answer important questions about the equilibrating tendencies of markets, but, since the preferences are artificial (chosen by the experimenter), they can tell us little about economies generated by actual preferences or the predictive power of these preferences for market outcomes.

Our paper connects the experimental market literature to an important strand of a very different experimental literature, measuring individual preferences. In that strand, subjects choose bundles of goods from each of a carefully designed collection of budget sets. Researchers typically find a high rate of consistency (e.g., with GARP, the generalized axiom of revealed preference) at the individual level, but also find striking diversity across individual preferences (e.g. for risk) in the subject population. However, that literature so far has said little about the aggregate economic implications of such preferences.

Connecting these two strands of literature allows us to address two fundamental questions that until now have been largely inaccessible. First, are revealed individual preferences typically consistent with the well-behaved (“nice”) sort of economic equilibrium usually considered in economic theory and deployed in applied work, or do they, instead, imply the sorts of predictive pathologies (non-existence, multiplicity, instability of equilibrium) highlighted in SMD? Second, are individual preferences sufficiently stable (over time and context) for the equilibrium predictions (whether nice or pathological) to accurately describe subsequent market outcomes? Answers to these ques-

tions are crucial for assessing the usefulness of both general equilibrium theory and of preferences measured in the lab.

Our experiment focuses on risk preferences (preferences for Arrow securities) and our markets are therefore financial (Arrow securities) markets. We chose this domain largely because it is quite challenging for the two fundamental questions. For the first question, preferences for risk documented in e.g. Choi, Fisman, Gale and Kariv (2007) show high degrees of convexity and non-monotonicity that we expected to make SMD-like pathologies particularly likely. For the second, prior literature suggests that risk preferences might be particularly unstable. However, we emphasize that our questions and methods are quite general and should extend to preferences over, and markets for, many other sorts of goods.

In Phase 1 of our experiment, subjects choose bundles of two Arrow securities (x, y) on 25 separate exogenously provided budget lines. Subjects are randomly and permanently assigned either as x -types with fixed initial endowment $(\omega_x, \omega_y) = (100, 0)$ or as y -types endowed with $(0, 76)$, and the budget lines differ only in the relative price of the two securities. Subjects face each price twice in an erratic order, allowing us to measure the stability of their revealed preferences. Using a simple non-parametric algorithm, we classify subjects as having “high convexity” preferences (i.e., as Hx or Hy , depending on endowment type) or as “low convexity” (Lx or Ly).⁴ Phase 2 of the experiment allows actual market trade using the tatonnement institution in four economies that use the classifications as building blocks. In a $HxHy$ economy, the Hx types and Hy types can only trade with each other. We run markets in that economy simultaneously with a separate $LxLy$ economy market comprised of the other subjects. Finally, we re-partition subjects into $HxLy$ and

⁴A high convexity subject will tend to view x and y as not good substitutes (in our experiment, perhaps due to high risk aversion), and a low convexity subject will tend to view them as more substitutable (perhaps due to low risk aversion). We define “more convex than” formally in Online Appendix A.4, Definition 1.

$LxHy$ economies and run markets for them simultaneously.

We calculate theoretical general equilibrium for any economy, including those run in Phase 2, using only the constituent subjects' raw Phase 1 choices. The revealed excess demand function, $z(p)$, is the sum x of the x-components of those choices at each price p , less the x-component, ω_x , of the aggregate endowment. We have a "revealed competitive equilibrium" (RCE) price wherever this excess demand function crosses zero. We can also use Phase 1 choices to directly construct other standard general equilibrium objects such as Edgeworth boxes and offer curves, and so obtain predicted (RCE) allocations. These constructions enable us to address the first fundamental question, on whether revealed preferences imply nice or pathological economies. The market outcomes in Phase 2 address the second question, on the predictive power of RCE prices and allocations. For example, we show that RCE prices in $HxHy$ economies are systematically lower than in $LxLy$ economies, and we test such predictions against the observed tatonnement market outcomes.

The results are striking. In Phase 1, our individual subjects reveal preferences that can easily produce pathological economies. Many subjects reveal very convex and/or non-neoclassical preferences, and we show how these can combine to produce the sorts of pathologies described in the Sonnenschein-Mantel-Debreu theorem. Specifically, about half of the simplest (most homogeneous) economies we can create from our subjects' Phase 1 choices ("pair-replica" economies) are pathological, e.g., have non-existent or multiple equilibria, and thus do not yield determinate GE predictions. Nevertheless, pathologies hardly exist in the 12-agent, *heterogeneous* economies we construct for Phase 2 market trade using the (Hx, Hy, Lx, Ly) building blocks. In each of 4 sessions, each of the 4 relevant excess demands functions crosses zero exactly once and thus provides a nice, unique RCE prediction.

Next, we examine whether preferences and the economies they give rise to are stable enough to predict behavior in markets. We use the degree of preference stability observed in Phase 1 to classify the RCE predictions as either “robust” or “fragile” to fluctuations in preferences. We show that market outcomes in Phase 2 are generally consistent with RCE predictions — observed prices and allocations usually converge to RCE values, and all of the exceptions are for economies a priori classified as fragile.

The final part of our analysis investigates why simple economies are so often pathological yet all of our building block economies are nice, and most of them robustly so. We use subjects’ choice data again to suggest a simple answer: diversity. The heterogeneity in revealed preferences seen in our study (and in all previous studies of which we are aware) tends to cancel out the pathologies as we aggregate more and more different traders into the economy. Such taming by preference diversity was hypothesized in prior theoretical literature including Hildenbrand (1983), Grandmont (1987, 1992), Hildenbrand (1994), Giraud and Quah (2003), and Hildenbrand and Kneip (2005). Using a variant of an index defined in that literature, our study provides (to our knowledge) the first direct empirical evidence that actual preferences are heterogeneous enough to overcome aggregation pathologies.

The organization of our paper is a bit unusual in that we interweave theoretical predictions with empirical results; the goal is to minimize the need to backtrack or skip ahead on first reading. Section 1.1 reviews relevant literature and Section 2 lays out the experimental design. Section 3 notes some theoretical implications of preference rationality and convexity, and checks their consistency with our Phase 1 data. Section 4 tackles the first fundamental question, whether revealed preferences imply nice or pathological economies. It shows how we use Phase 1 preference measurement to generate general equilibrium predictions (RCE) of price and allocation and then

presents the Phase 1 empirical results. Section 5 tackles our second question, on the predictive accuracy of RCE. It shows how we use variation in Phase 1 preference measurements to predict robustness and fragility, and then compares these predictions to Phase 2 market outcomes. Section 6 discusses aggregation and diversity, and Section 7 summarizes, comments, and points to possible future research. Details of algorithms (e.g., on sorting subjects) are collected in Online Appendix A.1, A.2, and A.3, and theoretical support for these and other procedures is relegated to Online Appendix A.4. Online Appendix 8.5 and 8.6 collect supplementary data visualizations, and Online Appendix 8.7 is a copy of instructions to subjects.

1.1 Related Literature

Our paper builds on the extensive “experimental revealed preference” literature, which studies how subjects make choices on budget lines. Prominent examples include Andreoni and Miller (2002) and Fisman, Kariv and Markovits (2007), who study other-regarding preferences, and Ahn, Choi, Gale and Kariv (2014) who study preferences for ambiguity. Choi, Fisman, Gale and Kariv (2007), Hammond and Traub (2012) and Halevy et al. (2018) study revealed preferences over risky assets, using designs related to Phase 1 of our experiment. An important difference in our study relative to this literature is that we fix subjects’ endowments and vary prices in order to study the economies that arise from subjects’ preferences. Most of this literature by contrast changes both endowment and price in order to test subjects’ consistency e.g., with GARP.

Our work is also connected to a rich literature stretching back to Chamberlin (1948) and Smith (1962) studying market behavior in *induced preference* economies. In the simplest of these, buyers and sellers are given marginal costs and reservation values (which together form economies with clear competitive equilibria), form prices in interactive markets and earn money based on the

distance between transaction prices and costs/values for traded units (thereby inducing the relevant preferences). These markets show a remarkable tendency under many institutions to produce competitive equilibrium prices and allocations even with relatively few (8-12) participants and little information. Williams et al. (2000) extend these methods to a general equilibrium setting in which subjects are given preferences over multiple goods by providing a matrix of cash payoffs for various terminal allocations and finds, again, remarkable equilibrating tendencies in competitive markets for economies with unique “nice” equilibria (see also Gjerstad (2013)). Our contribution is to study similar questions in settings in which economies derive from subjects’ own preferences rather than from experimenter-chosen preferences artificially induced via monetary payoff functions.

Relatedly, an experimental general equilibrium literature explores (again with induced preferences) market behavior in economies suffering from SMD-like pathologies like global instability and multiplicity. For instance Plott (2000) studies two-good economies with “backwards-bending” supply functions and multiple equilibria and shows that equilibrium selection in such economies are predicted by classical Walrasian disequilibrium dynamics (see also Plott and George (1992)). Anderson et al. (2004) and Gillen et al. (2020) study Scarf economies with no stable competitive equilibria and show evidence of non-converging price orbits and cycles instead of convergence. Crockett et al. (2011) study economies with multiple extreme (highly payoff inequitable) competitive equilibria and show that Walrasian dynamics combined with initial conditions strongly predict price trajectories and allocations. Goeree and Lindsay (2016) study how market design features can influence market performance in the face of such pathologies. This literature imposes economies with pathologies using induced preferences and studies resulting behavior; by contrast we study whether and to what degree such pathologies arise from actual preferences.

Because our experiment involves risky assets, it also relates to a growing experimental asset

pricing literature studying aggregate market behavior in which subjects supply their own intrinsic preferences over risky assets. One prominent strand of this literature tests broad predictions of the capital asset pricing model (e.g., portfolio separation and mean-variance efficiency) by having subjects trade multiple risky assets together with a riskless asset (e.g., Levy, 1997; Asparouhova, Bossaerts and Plott, 2003; Bossaerts and Plott, 2004; Bossaerts, Plott and Zame, 2007). A more distantly related strand studies asset bubble formation in long-lived assets, e.g., Smith, Suchanek and Williams (1988); Asparouhova, Bossaerts, Roy and Zame (2016), Breaban and Noussair (2015) and Crockett, Duffy and Izhakian (2019)). Importantly, individual preferences are not measured and aggregated to form market predictions, although in several papers a simple measure of risk preferences is implemented following the market experiment to be used as a covariate with individual or market behavior.

Two experimental finance papers report findings most closely related to ours. Bossaerts, Plott and Zame (2007) study experimental CAPM markets and show that aggregate outcomes in these markets are much better organized by theoretical predictions than is individual behavior (see also Bossaerts et al., 2010). Our results show a related contrast between individual vs. aggregate demand. Also, Biais, Mariotti, Moinas and Pouget (2017) measure average individual demand for a risky asset elicited using both (i) a BDM-like procedure (one price is randomly selected for payment) and (ii) a procedure related to market clearing (the price selected for payment is the one that minimizes excess demand summed across participants). Their focus is on the impact on demand of exogenous changes in aggregate risk, but they also find only minimal discrepancies between demands elicited under the two elicitation procedures. In contrast we fix aggregate risk and instead focus on how market outcomes respond to exogenous changes (via sorting) in revealed

preference convexity.^{5,6}

2 Experimental Design

Every session of our experiment consists of 28 subjects who participate in two successive Phases. As detailed below, in Phase 1 we measure subjects' preferences by presenting them with dozens of budget lines. Then we classify subjects according to the convexity of their revealed preferences and use these classifications to sort subjects into 12 person market economies (as we note below 4 subjects in each session do not participate in these markets). Finally, in Phase 2 each subject participates in two consecutive tatonnement markets for those economies (Section 2.3).

2.1 Eliciting Preferences

In Phase 1, subjects are endowed with an initial endowment $\omega = (\omega_x, \omega_y)$ of Arrow securities for two equally likely states, X and Y , and choose allocations (x, y) on budget lines $px + y = p\omega_x + \omega_y$ generated by 25 distinct exogenous prices p (normalized as the price of x with y as the numeraire). Subjects earn x (resp. y) tokens if X (resp. Y) occurs. At the beginning of the experiment, subjects are randomly assigned an *endowment type* which they maintain over all decisions (and both phases) of the experiment. The 14 x-type subjects have endowment $(100, 0)$ at each of their

⁵More distantly related is Buddish and Kessler (2018) who also study the relationship between separately elicited preferences and institutional performance. They assess the performance of a combinatorial assignment mechanism for class schedules by comparing the schedules subjects submit to the mechanism to the preferences they declare in a separate, unincentized task.

⁶Because our interest is in whether individual behavior translates to markets, also relevant is a small literature on the way preferences commonly measured in the laboratory express themselves differently in the context of naturally occurring markets – see Harrison et al. (2007) (risk), List (2006) (social preferences), and List (2003) (the endowment effect).

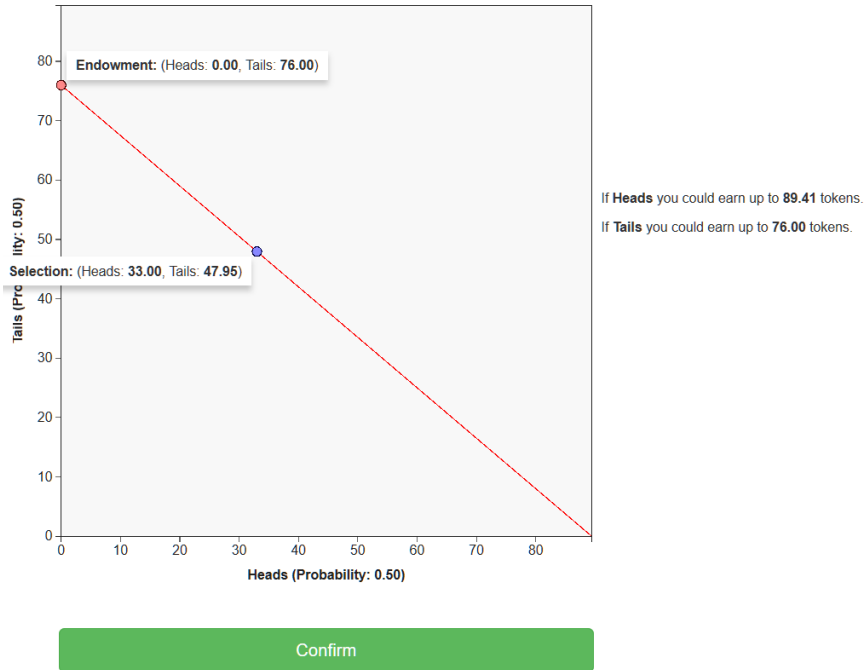


Figure 1: User interface. States X and Y are referred to as Heads and Tails respectively. Endowment here is $\omega = (0, 76)$, price is $p = 0.85$ and intercept values are shown in bold on the right. The user clicks any point (x, y) on budget line, e.g., $(33.00, 47.96)$, to check contingent payoffs, and clicks the green Confirm bar to make it her final choice.

decisions and the 14 y-type subjects have endowment $(0, 76)$.

Figure 1 shows a typical decision problem for a y-type subject; the red dot represents her endowment. The subject can click any point on the budget line to see a text box displaying the corresponding (x, y) values; the last point clicked becomes her actual choice when she clicks the Confirm bar (or when a countdown clock expires);⁷ subjects who never click simply receive their endowment for the period. The realized state is determined only at the end of the session.

⁷The clock expired at 120 seconds for choices 1-2 and the expiration time gradually decreased down to 40 seconds for the 23rd choice and beyond. Whenever all subjects confirmed their choices before expiration, the next choice was presented immediately.

Phase 1 is divided into two parts, each consisting of 25 choices. In Phase 1P (for “practice”) subjects make decisions on budget lines generated by 25 prices between 0.2 and 5 presented in an erratic order. The sequence presented is always (1.06, 0.36, 1.31, 0.64, 5.00, 2.00, 0.70, 1.43, 2.81, 0.57, 0.83, 0.28, 2.33, 1.00, 0.43, 0.89, 1.57, 1.21, 0.94, 0.50, 3.57, 0.76, 0.20, 1.75, 1.13). In Phase 1R (for “revealed”) each subject again faces the same price sequence.^{8,9}

2.2 Sorting Subjects

At the end of Phase 1 (after period 50) the computer assigns each subject to one of three *preference types*, H, L, or E, based on her 1R choices. As detailed in Online Appendix A.1, group E consists of the two subjects of each endowment type who made the most erratic (or “noisy”) decisions in terms of cumulative deviations from monotonicity.¹⁰ For each of the remaining 24 subjects, we compute a statistic, S_x , the sum of the x -components in their chosen allocations in Phase 1R at prices $p = 0.76$ and at the two adjoining prices, 0.70 and 0.83. Section 3.1 below shows that S_x is a non-parametric local measure of preference convexity. Of the 12 x -endowed (resp., y -endowed) subjects, the six with lowest S_x are assigned to group Hx (resp., Hy). The remaining six subjects of each endowment type are assigned to the low convexity groups Lx and Ly.

In each session, the Hx, Lx, Hy, and Ly groups are the building blocks for constructing four

⁸Prices are not shown explicitly and nothing special happens between periods 25 and 26. Thus, given the erratic price sequence, we do not expect subjects to realize that the sequence is repeated.

⁹To our knowledge, ours is the first experiment in the experimental revealed preferences literature to elicit preferences on every budget line twice. Doing this provides us a novel way of evaluating the consistency of elicited preferences, and may be of independent value in future work.

¹⁰Specifically, the software calculated, for each subject, a “noise measure” consisting of the sum of the changes in demand (across prices) that were accompanied by changes in the sign of excess demand. The 2 subjects from each endowment type with the largest such measure were classified as “E” by the software. An alternative measure we might have used is the distance between 1P and 1R, but this turns out to be highly correlated with our noise measure.

distinct 12-subject economies to be studied in Phase 2. The “Main” economies (Phase 2M) are HxHy — the 6 most convex x-endowed subjects together with the 6 most convex y-endowed in that session — and its opposite, LxLy. The two “Shuffle” economies (Phase 2S) re-sort the subjects into HxLy, which matches the 6 revealed most convex of the natural suppliers of x with the 6 least convex of the natural buyers of x, and LxHy, which does the reverse. Each subject of preference type E is essentially a “non-voting” participant in Phase 2 markets: her choices (and endowments) are ignored by the mechanism when it computes excess demand Z .¹¹

2.3 Running Markets

In Phase 2, we first run the two Main markets in parallel, and then run the two Shuffle markets in parallel. In each of these markets, subjects trade using a version of the venerable Tatonnement adjustment process (Walras, 1877).¹² Our tatonnement implementation is similar to that used by Guler et al. (2018) in an induced value experiment. In each round of tatonnement, a price is announced, and each subject selects an allocation along her corresponding budget line (for x-types the x-intercept is $x = 100$ and for y-types the y-intercept is $y = 76$), using essentially the same interface as in Figure 1. The algorithm computes excess demand given these choices, and

¹¹We chose a priori to exclude these four ex ante most confused subjects from the market in programming our software, but in retrospect it seems that most E subjects actually made fairly stable choices and this design decision had little effect on the results. Future researchers interested in replicating or extending our approach may wish to explore improvements on our filtering rule, or may simply elect to keep all subjects in the market.

¹²We chose the tatonnement institution for three reasons. First, it requires minimal additional training for subjects who have completed Phase 1. Second, unlike popular alternatives such as the continuous double auction, under tatonnement traders’ allocations do not change until the market converges. Hence the general equilibrium predictions (described below in Section 4.2) need not be updated while the market is in session. Third, in practice the tatonnement institution has proved fairly effective at producing competitive outcomes in induced preference market experiments (e.g. Guler et al. (2018), Joyce (1984), Bronfman et al. (1996)).

implements a price change with the same sign as aggregate excess demand. The process is repeated until the process has converged in either excess demand or price (or both).

The initial price in round 0 is always 1.57. Excess demand in round $t \geq 0$ is $Z = D - S$, where $S = 6(100) + 6(0) = 600$ is the total endowed supply of good x , and D is the sum over the 12 market participants of the x -components selected in that round. The tatonnement algorithm selects a price increment for round $t + 1$ that is a prespecified monotone and sign-preserving function of Z . For example, if $Z < 0$ in round t (as we anticipate at $p_0 = 1.57$ given the relative scarcity of security y), then the price p_{t+1} in the next round is strictly less than the current price p_t . Price changes in radians are proportional to Z , with the proportionality constant decreasing in later rounds. The adjustment process is deemed complete, and trades are consummated, when the absolute value of per capita excess demand $|Z|/12$, is less than 2.0 for three consecutive rounds or, if that criterion is never met, after the 25th round. Subjects receive their actual demands in the final round of the adjustment process, with the (in practice, very small) difference between the sum of these demands and the aggregate endowment being supplied or consumed by the experimenter.¹³ See Online Appendix A.2 for complete details and discussion of alternative versions.

2.4 Implementation Details

At the conclusion of Phase 2 we pay subjects for their choice in one period determined randomly as follows. Each subject draws one ball (with replacement) from a bingo cage filled with one hundred balls numbered 1-100. A draw between 1 and 50 means payment for chosen allocation in the Phase

¹³Other variations on this mechanism are easily imaginable and could be investigated in future research. For instance, one can simply give subjects their initial endowments (no trade) in markets that failed to quantity-converge by round 25 (instead of consummating trade at the round 25 price). Exploring variations in the market mechanism (including more realistic mechanisms like double auctions) is an important task for future work.

1 period with that number. A draw between 51 and 75 means payment for post-trade allocation in the final round in Phase 2M, while a draw between 76 and 100 means payment for the post-trade allocation in Phase 2S. Thus Phases 1 and 2 have an equal probability of being selected for payment, as does each period within a given phase. After determining the relevant allocation (x, y) , the subject flips a coin. Heads (Tails) yields the payment $\$/3$ ($\$/3$), that is, one dollar for every three tokens allocated to security X (Y) in the relevant period.

We conducted four sessions between May and September 2016 at the Subotnick Financial Services Center Lab at CUNY Baruch, each with 28 subjects recruited via e-mail through the subject pool maintained by the Department of Economics and Finance. Including instructions, practice periods and payments, each session lasted about 120 minutes. Including the \$7 showup fee, the average subject earned \$24.08 with standard deviation of \$14.51.

3 Revealed Preferences

The first part of our analysis concerns a familiar sort of question: what do subjects' Phase 1 choices reveal about their individual preferences? Unlike most previous studies, our GE focus leads us to emphasize a particular aspect of those preferences: the revealed excess demand functions.

3.1 Predictions

In General Equilibrium theory, a pure exchange economy consists of a set of agents $h \in H$, each characterized by her endowments ω^h and preferences. Excess demand functions $z^h(p)$ are computed for each agent h and summed to form the aggregate excess demand $z(p) = \sum_{h \in H} z^h(p)$. *Competitive equilibrium* is defined as a price vector p^* for which $z(p^*) = 0$, together with the agents' post-trade allocations $\omega^h + z^h(p^*)$. Thus for any pure exchange GE economy, agents' preferences matter only

insofar as they affect their excess demand functions.

In our design excess demand can be read directly from subjects' choices, since we assign budget lines that rotate around a fixed corner endowment. Although this design choice allows us to observe excess demand without making parametric assumptions, such budget lines do not have interior intersections and therefore preclude the GARP-style rationality tests pursued in previous papers. On the other hand, unlike the prior literature, we assign subjects each budget line twice (in Phase 1P and 1R), allowing us a different and novel WARP-style rationality measure (the distance between 1P and 1R choices).¹⁴ We make use of this measure in Section 5.1 below to understand how preference consistency influences the predictive power of GE theory.

Here we are interested in a different sort of rationality, relevant for understanding the nature of excess demands and the economies they give rise to. Proposition 1 in Online Appendix A.4 shows that if an agent is “minimally rational” — i.e., her preferences are monotone (she prefers a larger payoff in each state), symmetric (label-independent for the two equally likely states) and continuous on (x, y) — then she will never choose less of the cheaper Arrow security than of the more expensive security. Thus we have¹⁵

Prediction 1. *At every price vector $p = p_x$, every subject's choice (x, y) will satisfy $x/(x+y) > 0.5$ (resp. < 0.5) if $\ln p < 0$ (resp. > 0).*

Online Appendix A.4 shows that this prediction arises from co-monotonicity of prices and

¹⁴WARP (the Weak Axiom of Revealed Preference) asserts that if bundle a is revealed preferred to b , then b is not revealed preferred to a . Thus a subject making two different choices in 1P and 1R on the same budget line violates WARP, and the distance between the two choices is a measure of the size of the violation, akin to the CCEI index deployed in Choi et al. (2007) to measure the size of GARP violations.

¹⁵We state this prediction using x-shares and log price to align with the conventions of the next subsection, where we will report relevant evidence.

chosen bundles, and that in our setting, it is also a consequence of first order stochastic dominance. However, the prediction does not require that preferences are neoclassical, much less consistent with expected utility maximization (i.e., representable via a Bernoulli function). Minimal rationality is sufficient.

The other characteristic of preferences we are interested in is *convexity*, which is particularly important for determining the nature of economies in the aggregate. Definition 1 of Online Appendix A.4 provides a partial ordering of preferences across agents by convexity, i.e., how fast the marginal rate of substitution changes along an indifference curve. Lemma 1 shows that, in our Arrow securities setting, this partial ordering corresponds to risk aversion. More importantly, Proposition 2 and its corollary show that agents with more convex preferences choose points on the budget line closer to its intersection with the diagonal $x = y$.

We now can justify using S_x (the sum of x demands at prices at or around 0.76, described in Section 2.2) to measure convexity and as a means to classify subjects. At price $p = 0.76$, all our subjects, whether endowed at $\omega = (100, 0)$ or $\omega = (0, 76)$, face exactly the same budget line. Since $p = 0.76 < 1$, all minimally rational agents will pick a point on the budget line below the diagonal by Proposition 1. By Proposition 2 and its corollary, more convex agents will pick closer to the diagonal, i.e., they will choose smaller $x^h(p)$ at $p = 0.76$ and adjacent prices. Thus

Remark 1. *More convex agents (those who regard the two goods as more imperfect substitutes) will make choices generating lower values of S_x .*

Based on prior research (e.g. Choi, Fisman, Gale and Kariv (2007)) we expect to find not only many high-convexity subjects but also significant heterogeneity in convexity across subjects. As we show in Section 4, both of these patterns are of central importance for posing and answering our main motivating questions.

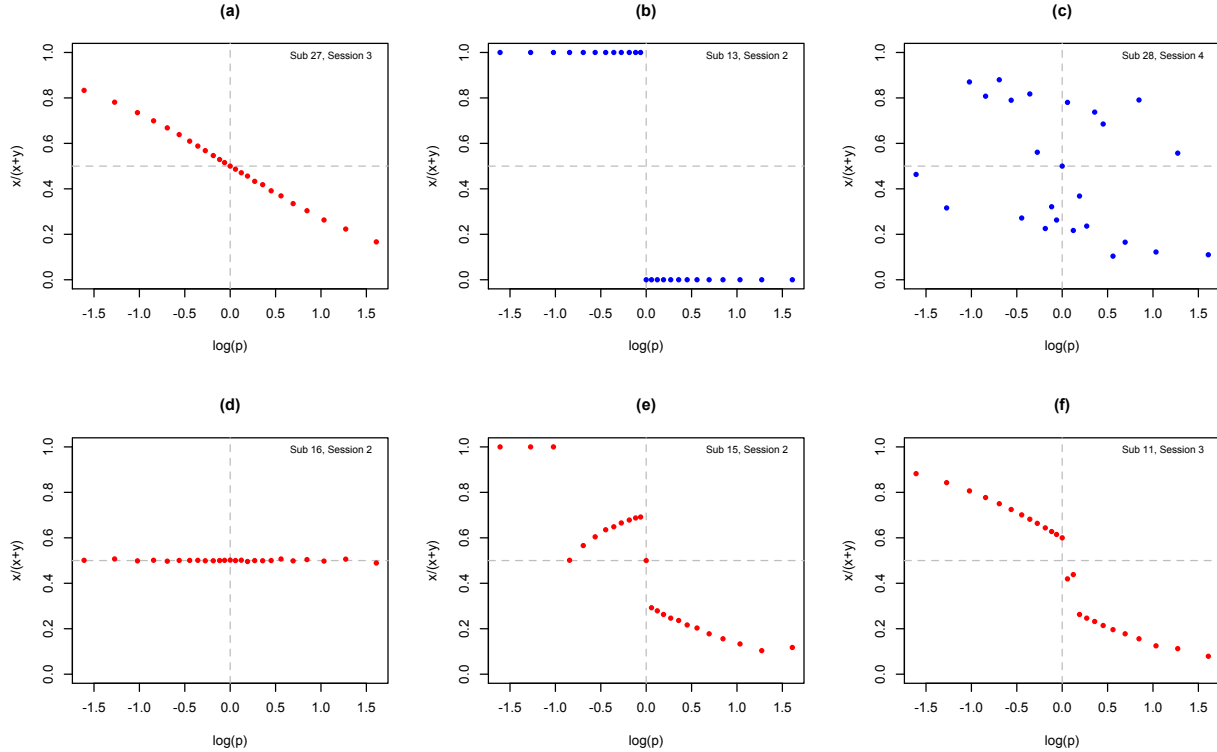


Figure 2: Phase 1R choices of selected individual subjects. Vertical axis is share of security X in the chosen bundle, and horizontal axis is log relative price of X .

3.2 Results

The preferences revealed in Phase 1 are similar to those in prior studies such as Choi et al. and, also as in prior studies, these preferences are highly heterogeneous. Figure 2 illustrates for 6 selected subjects, using the now-customary plots of choice share $x/(x + y)$ as a function of log price, for all 25 Phase 1R budget lines. About 7% of subjects make choices as in Panel (b), which can be described as treating the two goods as almost perfect substitutes or, equivalently, as having almost flat indifference curves, or as being almost risk neutral. The vast majority of our subjects treat the goods as imperfect substitutes (i.e. have convex preferences, are risk averse); Panel (a) is a particularly noiseless example. At the opposite end of the noise spectrum, the subject in Panel (c)

has choices all over the map including many that, contrary to Prediction 1, lie inside the upper right or lower left quadrants. Reassuringly, such behavior is atypical:

Result 1. *Only 5 of our 112 subjects violated Prediction 1 (chose dominated bundles) more than a third of the time, and overall fewer than 10% of choices violated that Prediction.*

Convexities are often quite extreme with subjects “hedging” by choosing a bundle very close to the diagonal $x = y$ for a range of prices around $p = 1$. Panel (d) shows an extreme example consistent with infinite risk aversion; more commonly, outside of a smaller price range, hedgers move towards (or even jump directly to) the less expensive corner.¹⁶ About 10% of our subjects make hedging-consistent decisions for more than 1/3 of prices. Panel (e) shows a weaker form of hedging that a working paper version of Choi et al. refers to as the “secure level heuristic.” For a range of prices, the subject purchases some fixed minimum quantity (here about 30) of the more expensive Arrow security and spends the rest of her wealth on the cheaper security. Nearly 30% of our subjects set a secure level that they enforce more than 1/3 of the time. Finally, panel (f) illustrates a heuristic we call “Ceiling,” adopted by 6% of our subjects: over a range of prices, pick a fixed maximum amount of the cheaper security and spend the rest on the more expensive security. Although these heuristics are inconsistent with standard neoclassical preferences, they do not violate minimal rationality. See Online Appendix A.5 for similar visualizations for all 112 subjects.

Figure 3 summarizes the characteristic of individual preferences that will matter most for our main research questions: the convexity of preferences as measured by S_x . The Figure plots empirical CDFs of this measure separately for x-type and y-type subjects. We note three important facts about these distributions. First, a randomization check: subjects assigned corner endowments

¹⁶As Choi et al. point out, such behavior is consistent with maximizing a utility function $U(x, y)$ with a kink on the diagonal, as in disappointment aversion (Gul, 1991).

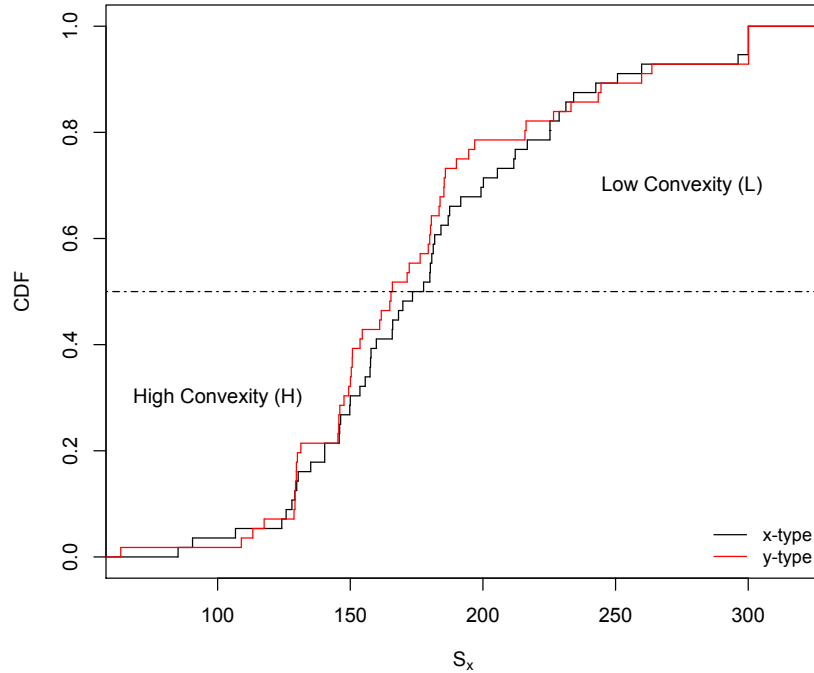


Figure 3: Cumulative density functions of convexity metric S_x for x-type and y-type subjects.

of x reveal no more or less convex preferences than those assigned y (Kolmogorov-Smirnov test, $p > 0.9$). Second, and more importantly, a significant proportion of subjects reveal substantial (local) preference convexity in their choices.¹⁷ Finally, convexities are not clumped around a few values, but rather spread over a wide range. Ninety-two percent of non-noise subjects with S_x lower than median (dotted line) are classified as Hx or Hy subjects, while 92% of non-noise subjects with S_x above the median are classified as Lx or Ly.¹⁸

¹⁷The minimum value of S_x for a minimally rational subject, corresponding to the “Highest” Convexity choice (i.e., $x = y$), is 129.4 (129.7) for endowment x (y) types. The maximum value of S_x , corresponding to the “Lowest” Convexity choice (i.e., choose only the cheaper asset, x), is 300 (300.1) for endowment x (y) types. The median subject who participates in Phase 2 markets (i.e., subjects typed as Hx, Hy, Lx, or Ly rather than E, as described in Section 2.2) has a value of S_x equal to 165.9; 79% of these subjects are closer to the minimum than maximum value.

¹⁸Of course, in practice, we sort subjects as H and L not relative to the entire sample as in the Figure but at the session level.

Result 2. *There is substantial heterogeneity in the convexity of preferences across subjects; the average Hx/Hy subject has considerably more (locally) convex preferences ($S_x = 141$) than the average Lx/Ly subject ($S_x = 212$).*

4 Revealed Economies

We now are ready to tackle our first fundamental question: what sort of economies arise from actual preferences? As noted earlier, in GE theory (pure exchange) *economies* are just collections of preferences (and endowments) that have calculable aggregate properties (e.g. equilibria), so our empirical question is: what types of economies are possible and common given the distribution of preferences our subjects reveal in Phase 1? Do Phase 1 choices produce nice economies with unique equilibrium predictions or pathological economies that have no clear predictions due to non-existence, multiplicity or dynamic instability of equilibrium?

4.1 Example Calculations

To address that question, we first illustrate how to derive competitive equilibrium predictions for any economy that is comprised of the preferences from a subset of our subjects. We reiterate that no structural or parametric assumptions, or estimation procedures, are needed: finding competitive equilibrium is a simple calculation from raw Phase 1 choices.

The calculation proceeds as follows. First, freely choose the set H of agents (subjects) that comprise the economy, and compile endowments $\omega^h = (100, 0)$ or $(0, 76)$ and actual Phase 1R choices $(x^h(p), y^h(p))$ for each p in the price grid for each $h \in H$. Next, calculate the individual excess demands $z^h(p) = x^h(p) - \omega_x^h$ at each price p , and sum them to get the aggregate excess demand function $z(p)$. A competitive equilibrium price occurs where $z(p) = 0$, or equivalently,

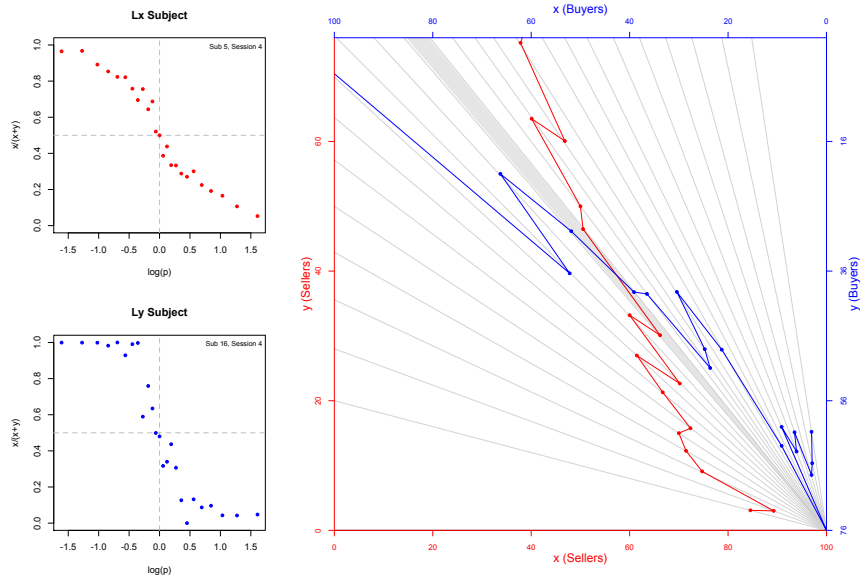


Figure 4: Phase IR choices for two subjects. Panels on the left display x -shares as a function of log price at grid points. Right panel is an Edgeworth box with the x -ended (y -ended) subject's empirical offer curve in red (blue); the RCE price interval is shown as a shaded wedge.

where aggregate demand equals the aggregate endowment. Since we observe $z(p)$ only at price grid points, our operational definition of revealed competitive equilibrium (RCE) is the interval between adjacent grid points such that z is positive at one point and negative at the other. RCE allocations are those arising from desired trades in the RCE price interval.¹⁹

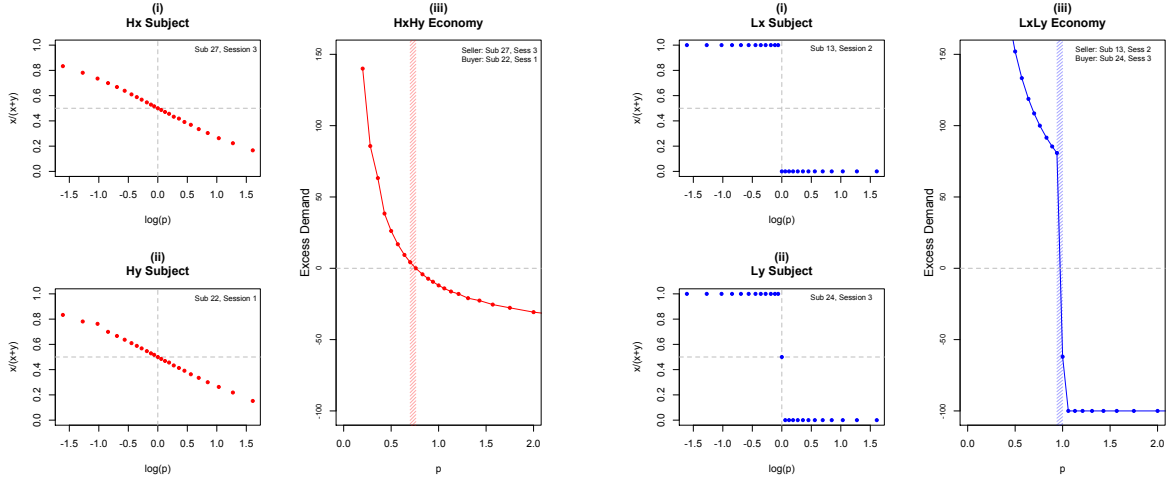
Figure 4 illustrates the calculations for an economy that consists of equal numbers of replicas

¹⁹For formalities, see Definition 2 of Online Appendix A.4. It may be worth remarking here that, by the intermediate value theorem, any interpolation (i.e., continuous extension of z from the price grid $\{0.20, \dots, 5.00\}$ to the interval $[0.20, 5.00]$) will satisfy $z(p) = 0$ for some p in the RCE interval. Our evidence is confined to the price grid, but presumably the aggregate complete excess demand function is closely approximated between grid points by standard interpolations, such as piecewise linear or cubic spline. If subjects are minimally rational (their preferences are continuous, monotonic, and symmetric), then strict convexity of preferences is a sufficient condition for the unknown complete z to be continuous and therefore have a root in the RCE price interval.

of a particular x-endowed subject (red) and a particular y-endowed subject (blue).²⁰ Their Phase 1R choices are displayed in the small panels on the left, and are re-plotted in the large Edgeworth box on the right. Here each price (no longer log price) defines a faint gray budget line emanating from the endowment point in the lower right corner of the box. Using the usual Edgeworth box conventions, the x-endowed (red) subject's chosen bundles (x, y) are plotted relative to the red axes' origin, and the y-endowed subject's bundles have coordinates shown on the blue axes with origin in the upper right corner. Excess demand $z(p)$ at any grid price p can be read as the difference between the x coordinates of the blue dot and the red dot on the ray corresponding to that price. Connecting the dots produces piecewise linear empirical offer curves. In this example, the offer curves have a unique intersection. The adjacent grid prices define the RCE price interval shown as a shaded wedge; somewhere in that price interval the two offer curves intersect so that aggregate demand equals aggregate supply and $z(p) = 0$. The RCE allocation box is the smallest axis-aligned rectangle containing the two blue and two red dots on that wedge.

Figure 5 shows streamlined examples of the RCE price calculation for two different pairs of subjects, by visualizing how individual excess demands give rise to aggregate excess demands. It illustrates how more convex revealed preferences tend, given our asymmetric endowments, to produce lower equilibrium prices. The pair-replica economy on the left consists of fairly convex subjects (classified as Hx and Hy in our experiment): their bundle shares respond only modestly as price moves away from 1.0. It takes a fairly low relative price, somewhere in the RCE interval $[0.70, 0.76]$, for aggregate demand $(x(p), y(p))$ to equal aggregate supply $(100, 76)$. The much less

²⁰We refer to the simplest (most homogenous) economies that can arise from a distribution of preferences as pair-replica economies, consisting of $N \geq 1$ clones of one x-endowed subject and N clones of one y-endowed subject. Using per-pair demands and endowments, Figure 4 does not change at all as N increases. In predicting competitive equilibrium outcomes, we implicitly assume that N is sufficiently large for attempts to exercise market power to be futile. Classic laboratory market experiments suggest that $N = 6$ is more than adequate.



(a) Two high convexity subjects.

(b) Two low convexity subjects.

Figure 5: Constructing excess demand and RCE price. Panels (iii) graph excess demand $z(p) = x_1(p) + x_2(p) - 100$ for the two subjects whose choices are shown in panels (i) and (ii) and whose x -endowments are respectively 100 and 0. The shaded vertical bar is shows the RCE price interval, which contains possible solutions of $z(p) = 0$.

convex subjects on the right (classified as Lx and Ly) are, of course, much more responsive to prices in the relevant range, and so they clear the market at prices near 1.0. Specifically, in this case the RCE price interval is $[0.94, 1]$.

4.2 Predictions

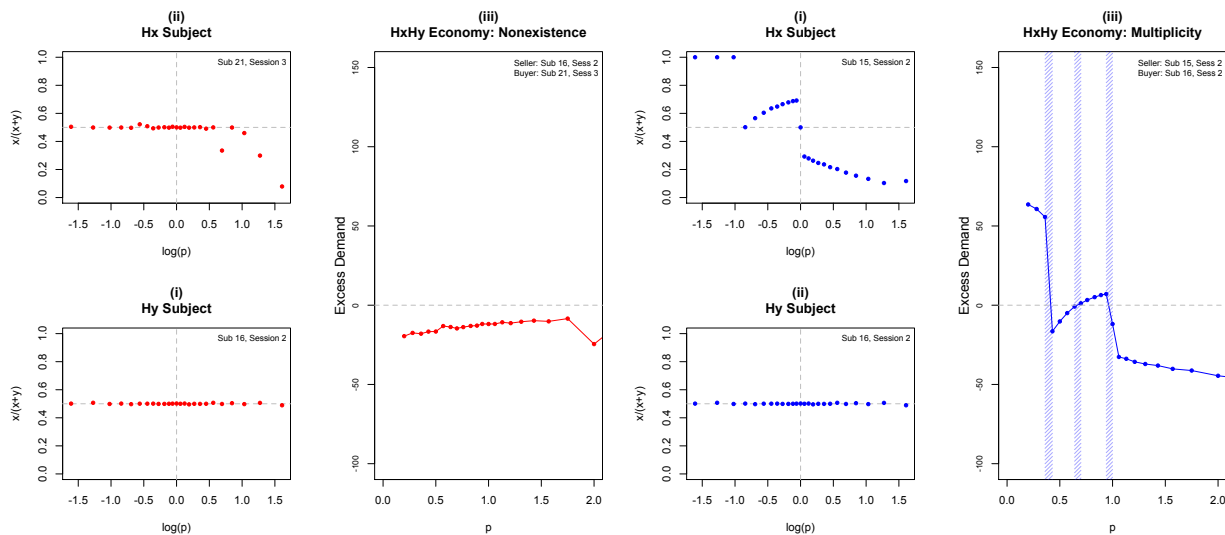
Proposition 3 in Online Appendix A.4 tells us that the insight from the last pair-replica example is quite general: in any pure exchange economy with greater aggregate endowment of x than of y , more convex preferences lead to lower equilibrium prices. The intuition is that the endowment asymmetry tends to push the price down, and the tendency is stronger with more convex preferences because they are less price-sensitive. By construction, Hx and Hy subjects have more convex preferences than the corresponding Lx and Ly subjects, so with asymmetric endowment $(100, 76)$ per agent pair, we have

Prediction 2. *In each session, the HxHy economy will have lowest RCE price and the LxLy will have the highest RCE price.*

Note also that the LxLy excess demands in Figure 5 are *steeper* in the neighborhood of the RCE than HxHy excess demands. We shall now see that this has implications for pathologies.

In the economies depicted in Figures 4 and 5, subjects' revealed preferences and demands produce a unique RCE, yielding a clear prediction of price and final allocation. The famous Sonnenschein-Mantel-Debreu theorem (SMD, sometimes described as the "anything goes theorem") suggests that we can not count on such pleasant situations. Sonnenschein (1973) showed that in a two-good economy such as ours, even nice neoclassical preferences can produce wild excess demand functions with multiple roots (and thus many competitive equilibria, some of which are dynamically unstable) or with no roots (and thus no equilibrium). Mantel (1974) and Debreu (1974) soon generalized that result to economies with more goods. Indeed, Mas-Colell (1977) showed that the set of general equilibrium prices is essentially arbitrary even if each consumer h has a nice utility function $U^h(x, y)$ representing continuous, monotone and strictly convex preferences. The SMD literature attributes these pathologies to "backward bending" excess demand due to preference "wealth effects" and very imperfect substitutes.²¹

²¹The intuition is as follows. You own a good X from which you derive a lot of your income. Suppose that the relative price of good X drops, so you now have less income. If goods are close substitutes, you simply keep more of the endowed good and buy less of everything else; no problem. But if you really like diversified bundles, then you still want to buy other goods W, Y etc, and that's going to cost you more of your X than before. So you may keep *less* X even though its price decreased. If your trading partners also regard X as not a good substitute for the other goods, then their demand for X may not increase enough to offset the reduction from people like you. Thus aggregate excess demand for X can, over some interval of prices, decrease as own price decreases, but for higher or lower prices outside this range it has the usual slope. This sort of non-monotonicity ("backward bend") is the key ingredient for the most famous examples of multiple equilibria in the GE literature (e.g., Scarf, 1960; Gale, 1963; Shapley and Shubik, 1977).



(a) Nonexistence.

(b) Multiplicity.

Figure 6: Examples of Pathological Pair Economies. Again, excess demand $z(p)$ in (iii) aggregates demand of the x -endowed and y -endowed subjects whose 1R choices are depicted in (i) and (ii).

To apply those ideas to our setting, we shall say that a particular economy is *nice* if (1) an RCE exists, (2) it is unique, and (3) it is “stable.” As detailed in Online Appendix A.4, stability means that the aggregate excess demand function is downward sloping at the RCE, as it is in both panels of Figure 5. Following the SMD literature, we shall say that the economy is *pathological* if it is not nice, i.e., if it has non-existent RCE or non-unique RCE or unstable RCE.

Our experiment was motivated in part by previous studies (e.g. Choi et al.) that find numerous subjects displaying extreme non-substitutability, suggesting vulnerability to SMD-like pathologies. Figure 6 shows examples, again using pairs of our own subjects. Subpanels i and ii of Panel (a) plot two subjects who exhibit extreme convexity (i.e., non-substitutability) resulting in non-existence of RCE prices over the observable range, as shown in subpanel (iii). Panel (b) includes (i) a subject who uses a fairly common non-monotonic strategy (observed also in the prior literature) and (ii) a highly convex subject, resulting in (iii) an economy with three distinct RCE, the middle one being

unstable. Since lower convexity subjects treat the goods as closer substitutes, and therefore are less likely to produce such pathologies, we have

Prediction 3. *“Pathological” equilibria will be most common in $HxHy$ economies and least common in $LxLy$ economies.*

4.3 Results

Figure 6 shows that pathologies are possible, but doesn't tell us whether they are typical. Do measured preferences tend to produce pathological economies, and so not allow general equilibrium theory to make determinate predictions? Because economies are aggregates, this is less a question about the nature of individual preference profiles, and more a question about the way such preferences are distributed *across* subjects. The most direct way to study the relevant distribution is to begin with the simplest (most homogeneous) economies that can arise from preference profiles in our sample: pair-replica economies.

We create a large and representative set of pair-replica economies by sampling repeatedly from the set of preferences measured in Phase 1. Each consists of equal numbers of replicas of (i) the revealed preferences of one of our 56 x-endowed subjects, randomly chosen and (ii) one of our 56 y-endowed subjects, randomly chosen. We compute the combined per-capita excess demand, and graph it in Figure 7(a). Pathologies are rampant, with excess demands failing to cross zero or (more frequently) crossing zero multiple times in about 40% of all the pair-replica economies we examine. Thus, the distribution of subjects' preferences is conducive to economies in which general equilibrium theory has no determinate predictions. We can also study the effect of convexity on the incidence of pathologies by sampling exclusively from Hx/Hy versus from exclusively Lx/Ly subjects. Panel (b) restricts the pairings to a Hx subject with a Hy , and panel (c) similarly restricts

to LxLy pair-replica economies (we will discuss the bottom two rows of this Figure in Section 6). Excess demands in panel (c) tend to be steeper around their crossing point and thus tend to cross zero only once, yielding a unique RCE, while HxHy pair replica excess demands are flatter and often cross zero multiple times or not at all. Overall, only 15% of pair-replica LxLy economies are pathological versus almost 60% of pair-replica HxHy economies, consistent with Prediction 3.

Result 3. *Not much more than half of pair-replica economies consisting of our subjects' revealed preferences have a unique competitive equilibrium. Pathologies (such as multiple or no equilibria) are much more common in high convexity than in low convexity pair-replica economies.*

The present exercise comes closest to examining individual preferences by studying maximally *homogeneous* economies, and the results show that there is nothing about the nature of these preferences themselves that protects against the types of pathologies cautioned in the SMD theorem. If economies constructed from our subjects' preferences turn out to be “nice,” it will not be because the preferences themselves are particularly nice. We can contrast this exercise with others that reintroduce the preference heterogeneity actually present in our data, by including more than two preference profiles in the economies we sample. This will allow us to study the further role that *heterogeneity* of preferences (as opposed to the nature of preferences themselves) plays in easing or exacerbating aggregation pathologies. Indeed, this is just what we do in constructing the economies we use for our markets in Phase 2. As explained in Section 2.2 above, for each of our 4 sessions our sorting procedure creates 4 economies, each consisting of six *different* pairs of subjects constituting the entire H or L pool for x-endowed and y-endowed subjects in a given session. Figure 8 plots the excess demand functions for all of these Phase 2 economies. Not surprisingly, we get nice excess demands for all 4 LxLy economies; each has a unique RCE price interval near $p = 1.0$. The 4 HxLy and the 4 LxHy economies are also nice, albeit with less steep excess demand functions. We were surprised to see that even the 4 HxHy economies turn out to be nice, although in some cases just

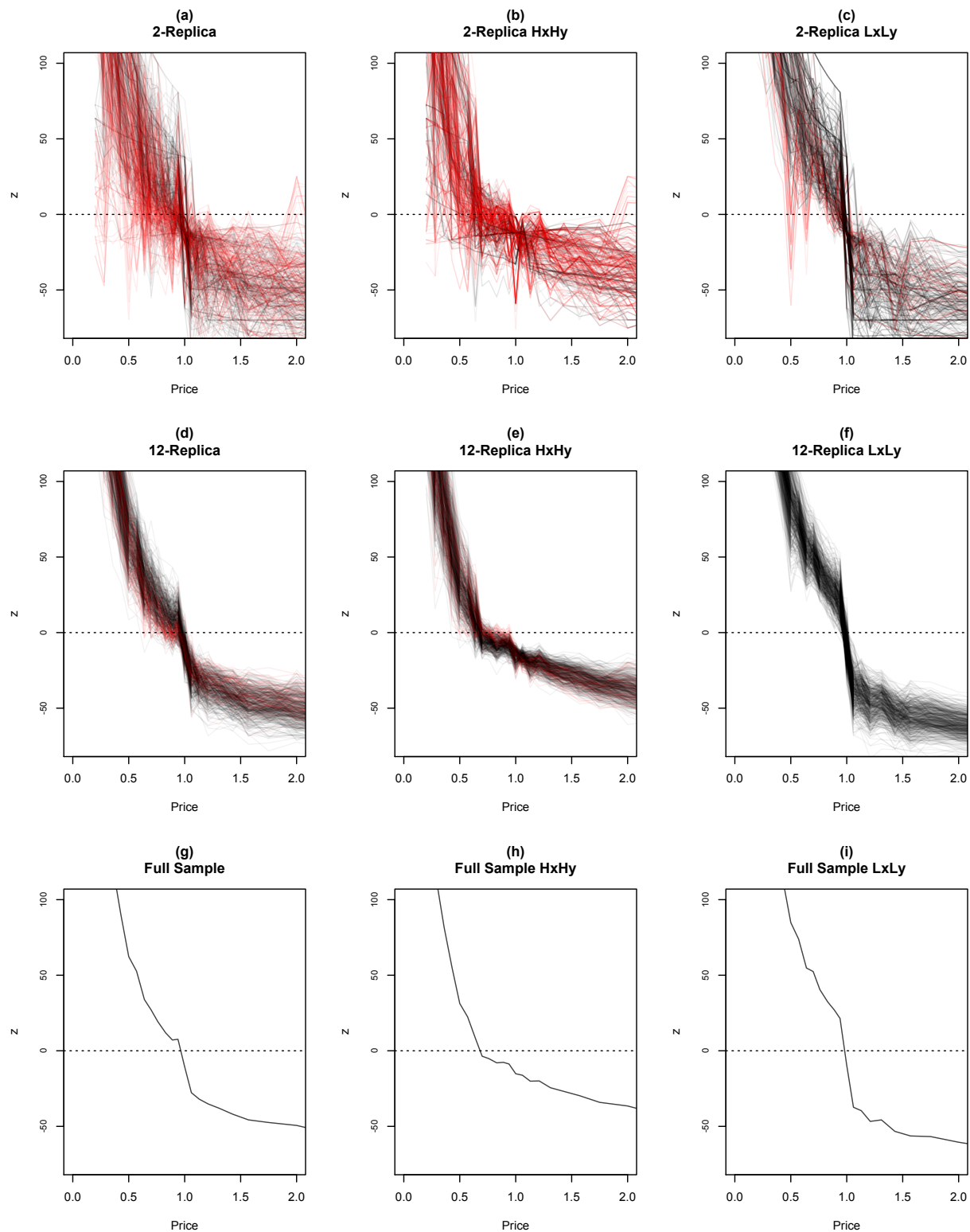


Figure 7: Thousands of pair-replica excess demand functions. Nice (pathological) excess demands shown in black (red).

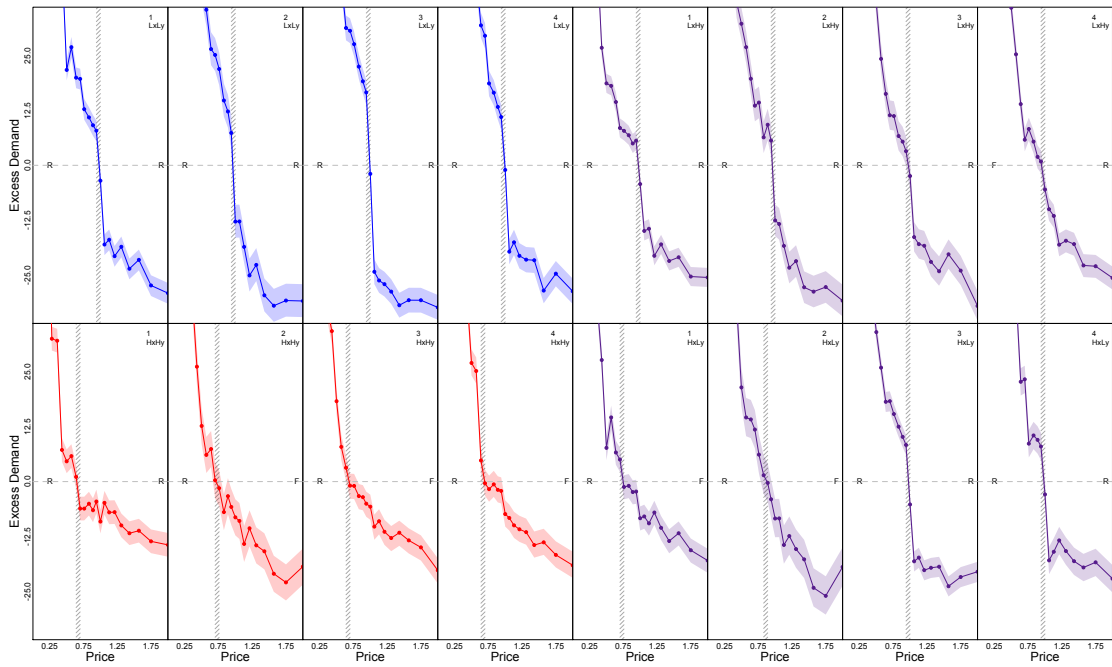


Figure 8: Phase 2 excess demand functions. Shaded vertical bars denote RCE prices. Shaded volatility bands and R[obust] and F[ragile] indicators are explained in Section 5.1.

barely so. As predicted, their RCE prices are always quite a bit below $p = 1.0$.

Result 4. *Of the 16 economies consisting of subjects sorted into Hx , Hy , Lx , and Ly blocks, none is pathological; each of them has a unique RCE.*

Why are pathologies so common in high convexity pair-replica economies but not in the 12-person economies we created for Phase 2 trade? In Section 6 we will investigate the apparent discrepancy between Results 3 and 4, and assess whether this is due mainly to luck or to systematic forces. To round out the present section, we note that inspection of Figure 8 discloses one more result that will be important for our Phase 2 markets:

Result 5. *Consistent with Prediction 2, in every session the Phase 2 $HxHy$ economy has the lowest RCE price and the $LxLy$ economy has the highest RCE price.*

5 Market Outcomes

We now turn to our second fundamental question: are preferences sufficiently stable for RCE prices and allocations (obtained from preferences revealed in Phase 1R) to reliably predict market outcomes observed in Phase 2? In our usage, preference stability (as opposed to RCE price stability discussed in the previous section) refers to stability of choice behavior over time (since Phase 2 comes after Phase 1) as well as over context (individual choice versus market behavior).

5.1 Predictions: Stability, Fragility and Robustness

We begin with stability over time. Recall that Phase 1 itself provides direct evidence since each subject faces the same budget line twice, first in Phase 1P and later in Phase 1R. Consider the decision error $d^h(p) = x_{1R}^h(p) - x_{1P}^h(p)$ between the x-components of subject h 's chosen Phase 1R

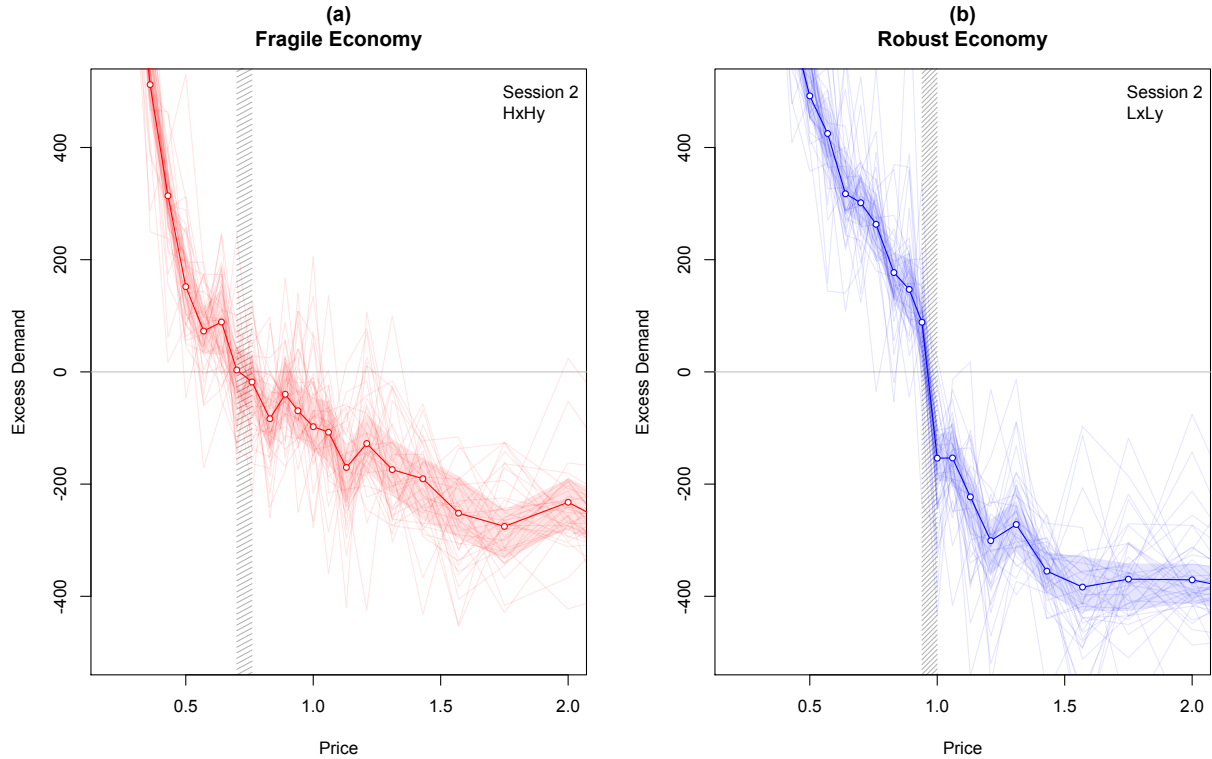


Figure 9: Two excess demand functions. Bold line is excess demand revealed in Phase 1R for the indicated 12 person economy. Faint lines are bootstrapped excess demands, and volatility band is shaded.

and 1P bundles. Online Appendix A.6 shows that most subjects make fairly (and often remarkably) consistent choices between 1P and 1R, but the decision error distribution has a long tail.

How much variability in an economy’s excess demand should we expect given those decision errors and to what degree should we expect this variability to interfere with equilibration? To find out, we construct *bootstrapped excess demands* as follows. For a given economy in a given session and each grid price p_k , begin with aggregate excess demand $z(p_k)$ and for each of the 12 active participants h , add or subtract (with equal probability) the absolute discrepancy $|d^h(p_j)|$ for a random grid price j drawn with replacement.

Figure 9 illustrates. It plots 12-person excess demands from our dataset (calculated directly

from Phase 1R choices) in a bold line and hundreds of bootstrapped excess demands in a paler shade. In panel (a), the bootstrapped excess demands frequently cross zero outside of the RCE. In such cases the tatonnement process can halt prior to convergence to RCE or even reverse course. We will therefore refer to economies of this sort as “fragile.” By contrast, in the economy in panel (b), excess demand is quite steep around the RCE and for this reason bootstrapped excess demands rarely lead to sign changes in excess demand. We classify such an economy as “robust” because instability in preferences over time seem unlikely to interfere with convergence to RCE. More precisely, we form “volatility bands” bounded by the 25th and 75th percentiles of thousands of bootstrapped excess demands at each price.²² An economy is defined as upside (resp. downside) *fragile* (F) if the volatility band crosses the horizontal axis $z = 0$ at any price above (resp below) the RCE interval; otherwise the economy is *robust* (R).

We predict that robustness and fragility will be influenced by our treatments. As discussed above, highly convex HxHy economies produce flatter excess demands than LxLy economies and therefore smaller (in absolute value) excess demands in the neighborhood of equilibrium. For any distribution of decision errors, preference fluctuations will create more sign changes in the former case, yielding the following

Prediction 4. *HxHy equilibria will be classified as fragile more often than LxLy equilibria.*

We hypothesize that markets are less likely to converge to equilibria that we can pre-classify as fragile because, in such cases, we expect fluctuations in excess demand to regularly interfere with the tatonnement process by altering the sign of excess demand. This gives us a final prediction:

²²Thus we implicitly use the interquartile range. An alternative, parametric, measure of volatility, the standard deviation of the discrepancy distribution, is less robust to outliers and so leads to a wider band. Nevertheless, this alternative approach (as well as several other variants we examined) leads to similar conclusions as those presented below.

Prediction 5. *Markets will be less likely to converge to fragile RCE than to robust RCE.*

5.2 Results

To evaluate Prediction 4, Figure 8 includes volatility bands for each of the Phase 2 economies and labels each upside (to the right of RCE) and downside (to the left) as either R or F. Inspection of that Figure yields

Result 6. *All our LxLy economies are robust, as are a majority of the shuffle economies, but most HxHy economies are upside fragile, consistent with Prediction 4.*

LxLy economies tend to more robust than HxHy economies mainly because they have steeper excess demand. It turns out that differences in volatility play little role; indeed, Online Appendix A.6 shows that decision errors for H subjects are very similar to those for L subjects. Thus again our sorting procedure generates an important treatment effect.

We now have direct predictions of Phase 2 market outcomes derived from Phase 1R revealed preferences, as well as second order predictions of their accuracy (or robustness) based on preference stability in Phase 1R-1P. Are these predictions useful, or are they perhaps undermined by the change of context from individual choice to market interaction?

As a preliminary, we note that although strategic behavior is possible in market institutions like tatonnement, there is no evidence of it in our 12-person Phase 2 economies. To show this we compare the demands subjects submit in Phase 2 markets to the choices they submit at the same prices in (strictly incentive compatible) Phase 1R elicitation.²³ Demands between these two

²³As described in Online Appendix A.2, our tatonnement algorithm assigns prices in Phase 2 on the same grid that was used in Phase 1, until doing so would produce identical prices two periods in a row, at which point prices are implemented on a continuum. Therefore, each subject faces several budget lines in Phase 2 that she faced (twice) in

contexts are not statistically different (Wilcoxon test $p = 0.82$). Moreover, the deviation between market choices and Phase 1R choice are not distributed any differently than the deviation between Phase 1R and 1P choices (Kolmogorov-Smirnov test, $p = 0.98$).²⁴ This suggests that subjects behave similarly in markets as they do in elicitation and therefore that our markets are large enough for strategic price manipulation to be negligible.

As a second preliminary, for all intents and purposes, all of our markets converged. Ten markets converged in excess demand (i.e. absolute per capital excess demands were less than 2) and were halted by the mechanism. The remaining 6 markets continued until the end of the period but had nonetheless effectively converged, with small absolute excess demands (under 5 per capita) and prices vibrating in a tiny range (under 0.1 on average) by the final three rounds.²⁵

Moving on to our main questions, Figure 10 plots predicted versus observed market behavior in two Phase 2M economies. The left side of each panel is essentially a Figure 8 panel with the axes flipped. RCE prices now are a horizontal band that extends through the right side of each panel, where we see the prices observed in the Phase 2 tatonnement market. In Panel (a) we have a robust economy, and the observed round-by-round prices quickly find the RCE band, vibrate around its top edge $p = 1.00$, and halt there in round 14. The economy in Panel (b) is upside fragile (F). The observed prices never quite reach the RCE band, and ultimately settle down at about 0.80, near the middle of the range where the volatility band overlaps the $z = 0$ line. Figure 11 shows similar Phase 1, which facilitates the direct comparison of choices between Phases 1R and 2.

²⁴In both of these tests, we avoid over-attributing independent variation to the data by taking the average deviation by subject and then conducting Wilcoxon and Kolmogorov-Smirnov tests on this sample.

²⁵The reason these markets were not formally halted by the mechanism is interesting methodologically for followup research. All but one of these markets generated prices extremely close to 1 and at this price risk-neutral subjects should be indifferent over all demands. Not surprisingly, then, almost all of the too-large excess demands in these markets were submitted by low-convexity Lx/Ly subjects.

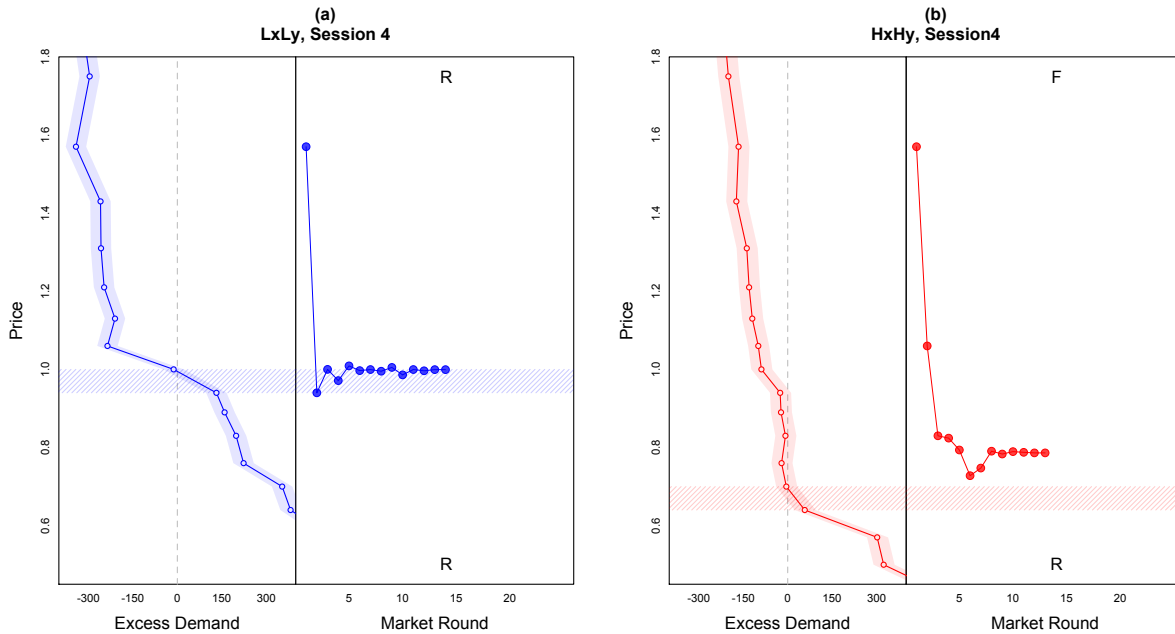


Figure 10: Price convergence in Phase 2M, Session 4. Left sides show inverse excess demand computed from Phase 1 behavior. Right sides show round-by-round tatonnement prices, and the designation as upside and downside Fragile (F) or Robust (R) at the top and bottom.

plots of market behavior in all 16 economies that we ran in Phase 2.

Despite the failure of the HxHy market in Figure 10 to converge to RCE, its final price $p \approx 0.80$ is still below the LxLy final price $p \approx 1.00$ in the same session, consistent with Prediction 2. The same is true overall; at $p < 0.001$ a paired Wilcoxon test rejects the null hypothesis that final prices are equal across pairs of economies in sessions and periods in which RCE differ. Phase 2 final prices thus support the directional comparative statics. Moreover, their numerical values are tightly correlated with the RCE predictions — the correlation coefficient is 0.85 ($p < 0.001$) overall.

Result 7. *Prices are strongly correlated with RCE predictions and within-session RCE price comparative statics are universally obeyed.*

What about the second order prediction (Prediction 5) — do RCE classified as robust (par-

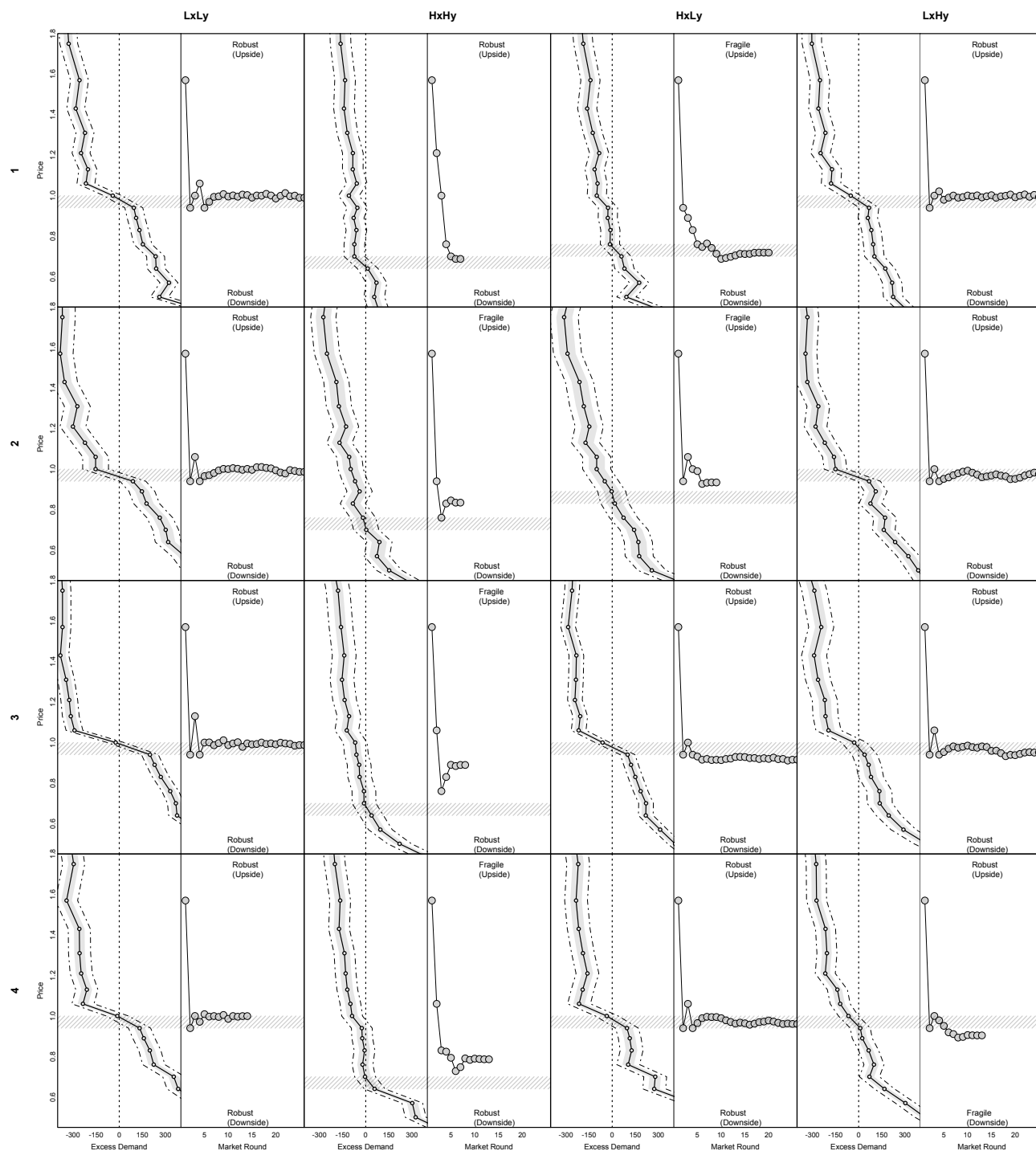


Figure 11: Inverse excess demand functions (left side of each panel) and time series of prices (right side of each panel) for each economy in the dataset. Panels are organized into columns by economy type and row by session (marked on the outside of the grid).

ticularly upside robust) actually predict final prices better than do fragile RCE? In all 11 of our markets with upside robust RCE, prices reach the equilibrium band and converge to a point at or below the band. In 10 markets, RCE are both upside and downside robust; in 9 of these prices not only reach the RCE but remain within its bounds. The only exception is HxLy in Session 3, where prices vibrate just below the tunnel. We conclude that when RCE are robust, markets reliably converge to the RCE. In 5 of our markets, RCE are upside fragile and in only one of these do prices reach as far as the competitive equilibrium. In one additional case, the RCE is downside fragile and in this case prices reach the RCE but continue to fall below the RCE bounds, ending at a lower price. Thus, markets typically fail to converge to fragile RCE.

Overall, robustness and fragility classifications have strong predictive power on convergence to the RCE. A Chi-square test allows us to reject the hypothesis that prices are equally likely to converge in fragile and robust economies ($p = 0.016$). The results also support the hypothesis that treatment interventions controlled the reliability of equilibrating dynamics by influencing the steepness of excess demand: prices reach the RCE in all of our LxLy economies (where our sorting algorithm is expected to give these the steepest excess demand functions), the majority of HxLy/LxHy economies, but only a minority of HxHy economies (expected to have the flattest excess demand functions). The price data thus suggest no additional instability due to market context vs individual choice.²⁶ More specifically, our data yield

Result 8. *Prices universally converge to robust RCE, but rarely to fragile RCE, supporting Prediction 5.*

Finally, we compare final allocations in Phase 2 markets to RCE predictions. Figure 12 plots mean offer curves for the 6 subjects of each endowment type for two example economies from our

²⁶Of those markets that terminate due to the mechanism halting, half (5 out of 10) converge to the RCE tunnel. Of those markets that continue slight fluctuations until the final round, 83% (5 out of 6) end in the RCE tunnel.

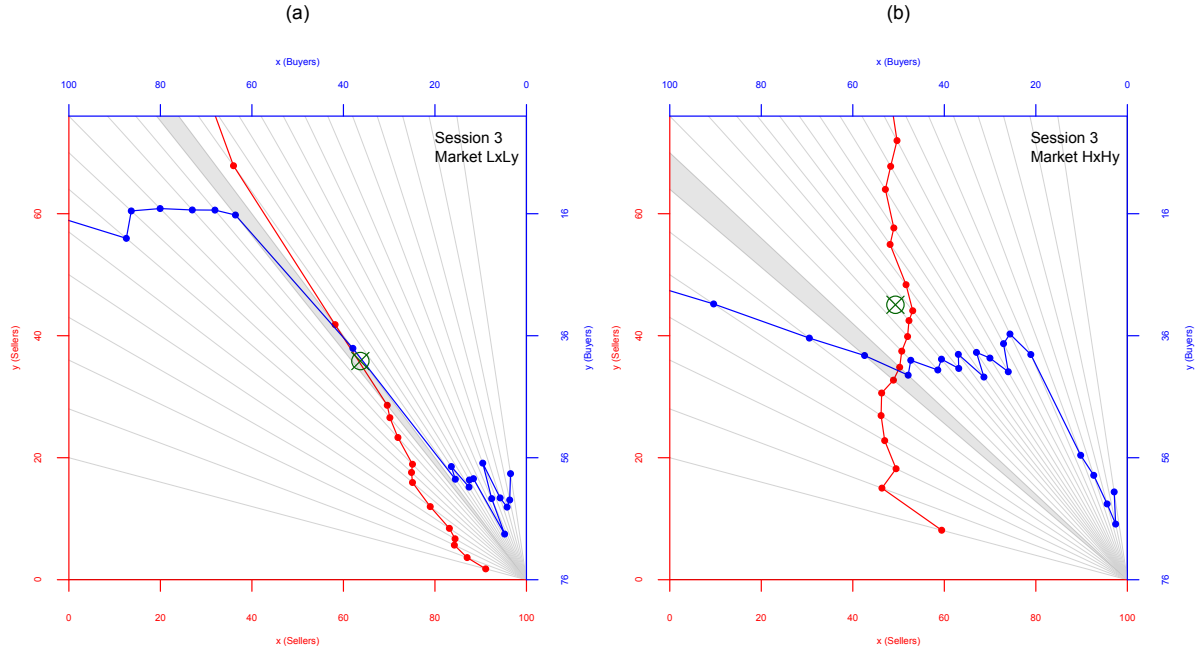


Figure 12: Two Empirical Edgeworth Boxes Mean offer curves and RCE price wedge drawn with the same conventions as in Figure 4; observed final allocation marked by circled green X.

data. The final allocation in the robust RCE depicted in Panel (a) is very close to the intersection of the two mean offer curves, and well within the RCE allocation box. The RCE is fragile in Panel (b) and the observed final allocation lies outside the RCE allocation box. Figure 13 shows similar figures for all 16 Phase 2 economies, with doubly shaded boxes indicating RCE allocations. In all, 75% of economies generate allocations inside the RCE boxes and all 4 exceptions have fragile RCE. This may reflect the predictive power of robustness, but it may also simply reflect the tendency of robust RCE to have larger allocation boxes.

Result 9. *Final allocations usually lie in the RCE allocation box, and always do so in economies with robust RCE.*

To summarize Phase 2 results, we find that (a) most markets (10 of 16 in price space, 12 of 16

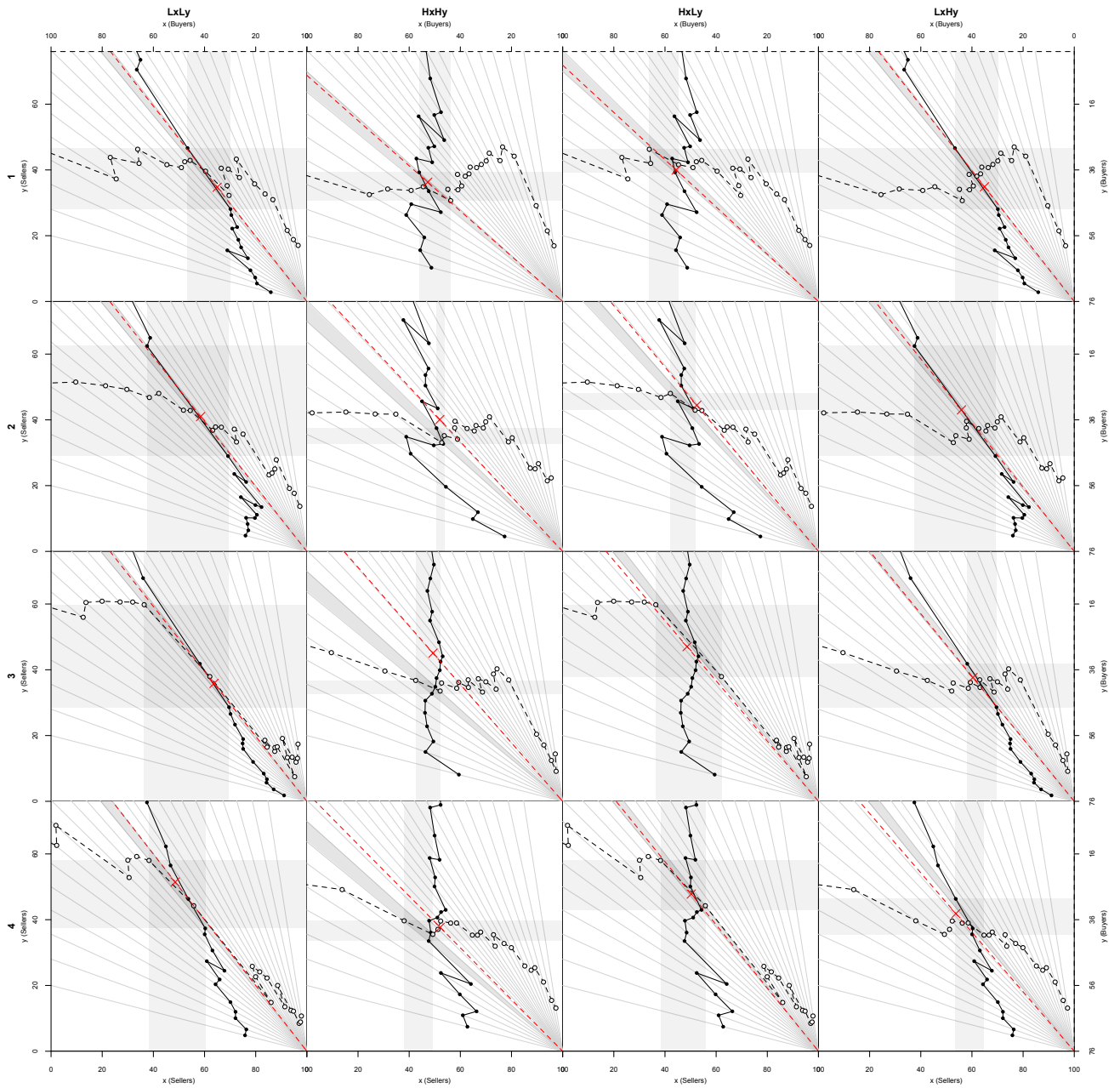


Figure 13: Empirical edgeworth boxes, plotting offer curves for the average subject of each type.

Panels are organized into columns by economy type and row by session (marked on the outside of the grid). Gray rays mark the price grid, shaded wedges mark RCE price bounds and shaded gray rectangles RCE allocation bounds.

Red dashed rays mark final prices from Phase 2 markets and red X's final allocations.

in allocation space) converge to the competitive equilibria implied by subjects' preferences revealed in Phase 1R, and moreover (b) decision errors in Phase 1, together with the shape of the excess demand function constructed from those revealed preference, allow us to predict which markets will fail to converge. Finally, preferences are stable enough that we can use subjects' own revealed preferences as a treatment instrument for causally influencing market outcomes.

6 How Diversity Cures Pathologies

One major puzzle remains. We saw that individual preferences revealed in Phase 1 are conducive to the sorts of pathologies (non-existence, multiplicity and dynamic instability of equilibrium) highlighted in the classic SMD literature. Why then does our procedure for building Phase 2 economies produce so many that are nice and robust?

We saw in Figures 6 and 7 that preference heterogeneity is part of the problem, but perhaps it is also part of the solution. The pair-replica economies featured there allow only two different revealed preferences. Our Phase 2 economies, by contrast, are built from the revealed preferences of 12 different subjects. Might that increase in diversity account for the difference?

Figure 7(d) shows excess demand functions for thousands of 12-agent economies constructed by drawing with replacement the Phase 1R choices of 6 x-endowed and 6 y-endowed subjects from the entire four session pool. The pathology rate in Panel (d) is only 15%, down from 40% for the corresponding pair-replica economies pictured in Figure 7(a). Panel (f) of Figure 7 shows that pathologies virtually disappear (the rate is less than 0.001%) when we impose the LxLy block restriction on our draws. The pathology rate is, of course, highest in the random HxHy economies shown in panel (e), but still is only about 15%. Aggregating our entire dataset into a single 112-agent economy as in panel (g), or into the 56-agent economies shown in (h) and (i), we find that

pathologies disappear entirely; all three large economies have robust and unique RCE.

These results suggest an empirical inversion of the Sonnenschein-Mantel-Debreu theorem. SMD shows that nice preferences can generate pathological excess demands in the aggregate; while our data, which include problematic revealed preferences (many of them very convex and/or not neo-classical), aggregate into nice excess demands, with orderly market outcomes, when we include enough subjects with differing preferences.

The idea that diversity might tame pathologies has a long history in general equilibrium theory, stretching back informally to at least Becker (1962). Formal analysis by Hildenbrand (1983) considers endowment heterogeneity, and work by Grandmont (1987, 1992), Hildenbrand (1994), and Giraud and Quah (2003), among others, consider preference heterogeneity. In particular, Hildenbrand and Kneip (2005) construct an index of behavioral diversity (IBD) that measures diversity across individuals' price sensitivity of expenditure shares. Since we have access to individual demand data, we are able to streamline their construction by looking directly at individuals' price sensitivity of demand. SIBD, our variant of their diversity index, is detailed in Online Appendix A.3.

The black line in Figure 14(a) plots the mean pathology rate in thousands of sampled economies as a function of their mean SIBD as we move from pair-replica (2-subject) economies to larger economies that match 6, 12, 24, 48 and 96 unique subjects sampled from our dataset of 112 subjects. The pathology rate indeed converges to zero as the SIBD increases. Results are similar for restricted matchings of only high convexity subjects (Hx and Hy, plotted in red), and only low convexity subjects (Lx and Ly, plotted in blue). Of course, the high (low) convexity economies have higher (lower) rates of pathology at identical values of SIBD than the unrestricted samples because they typically have flatter (steeper) excess demand functions. However, in all three cases,

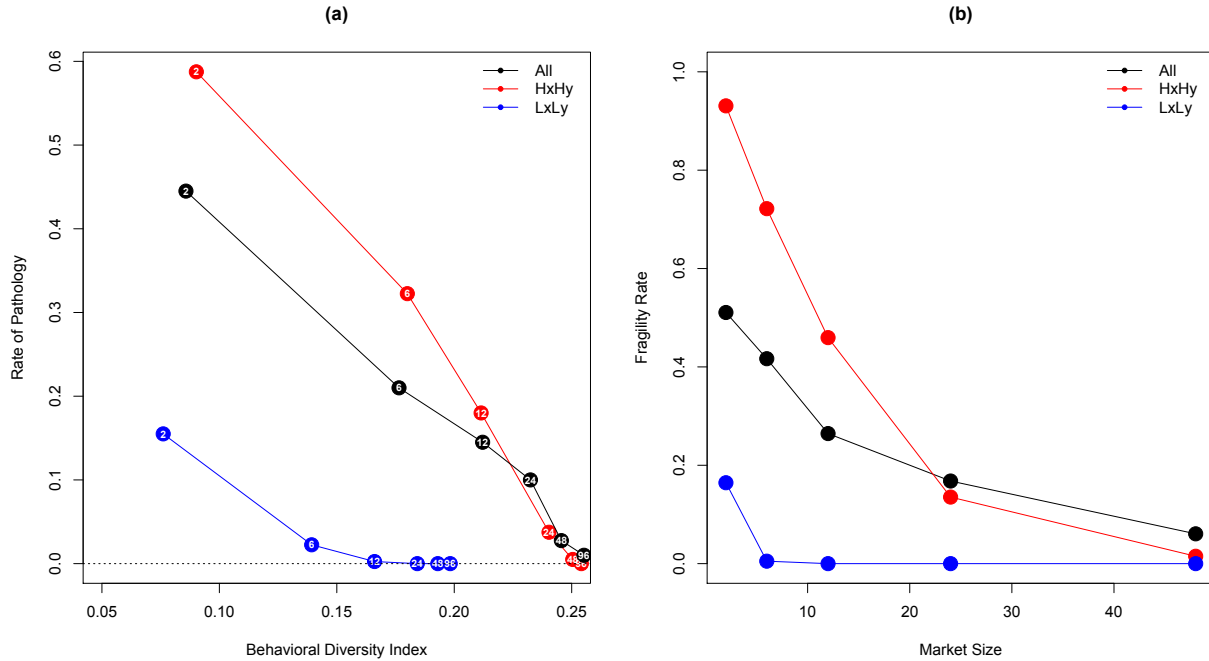


Figure 14: Market Diversity and Size Effects. Panel (a) shows rates of pathology as a function of behavioral diversity in the economy using simulations. Panel (b) shows rates of fragility as a function of market size.

pathology rates drop monotonically and eventually reach zero as diversity increases.

Aggregation removes the pathologies motivating our first fundamental question (Section 4); it also turns out to remove the effects of preference instability motivating our second question (Section 5). Panel (b) of Figure 14 plots fragility rates in economies built from individual excess demands of 2, 6, 12, 24, and 48 sampled subjects each. It shows how the fragility rate drops as market size increases.²⁷ This last result is a consequence of the well known Bienaymé formula, that variance of the mean shrinks linearly in the number of independent samples: it follows from the definitions that the per-pair volatility band shrinks and the economy is less likely to be classified fragile as we increase the number of sampled decision errors. In replica economies, of course, the volatility bands

²⁷To simplify computations in this exercise, we conduct this exercise using standard deviation volatility bands, and to simplify interpretation we plot only economies with unique equilibria.

would shrink as we increase the number N of replicas although SIBD would remain constant.

Aggregation thus improves the predictive power of general equilibrium theory in two distinct ways. First, revealed preferences, though often far from non-neoclassical, turn out to be so diverse that, once a sufficient number of different subjects join an economy, its aggregate excess demand function becomes nearly monotonic. Second, for standard statistical reasons, the impact of noise diminishes as we aggregate larger numbers of subjects. Consequently we can reliably predict general equilibrium outcomes using revealed preferences.

Result 10. *Pathologies — non-existence, instability, and multiplicity of equilibria — disappear in our data with increases in the number of participants with diverse revealed preferences. Fragility also disappears in larger economies due to attenuation of volatility.*

7 Conclusion

We introduce new experimental methods for studying the aggregate, general equilibrium implications of experimentally revealed preferences. Our experiment applies these methods to a particular but important sub-class: preferences for risky choice. Specifically, subjects reveal their preferences over bundles of two Arrow securities. We use their revealed preferences to sort the subjects into different economies with a range of general equilibrium predictions, and compare those predictions to subsequent market outcomes. These new methods allow us to address previously inaccessible empirical questions in general equilibrium and provide insight on the interpretation of preferences commonly measured in the laboratory.

What do we learn from our experiment? First, individual subjects' revealed preferences easily produce replica economies that suffer from the predictive pathologies highlighted by the Sonnenschein-Mantel-Debreu (SMD) theorem — multiplicity, non-existence, and instability of general equilib-

rium. At the individual level (and as in prior experimental work on revealed preferences), subjects reveal preferences that often suffer from non-substitutibilities and non-monotonicities so that replica economies formed from pairs of subjects suffer SMD pathologies (especially multiplicity and instability) almost half the time.

However, we find that preference heterogeneity across subjects has a powerful taming effect on aggregate excess demand as economies get even moderately large (and therefore diverse). Aggregating the choices of only 12 subjects produces nearly monotone excess demand functions that yield a unique competitive equilibria in each of the 16 laboratory economies we create. This is true even in the four economies consisting only of subjects with the most pathology-prone revealed preferences. Our empirical results reinforce the suggestion of some theorists, including Grandmont (1987) and Hildenbrand (1994), that human preferences indeed may be heterogeneous enough to overcome SMD pathologies.

Second, general equilibrium theory, applied to revealed preferences, predicts behavior in our markets quite well even though the preferences we study are not perfectly stable across decisions and our markets are not very large. Simply by partitioning subjects into economies according to their revealed preferences, we are able to influence the observed level of market prices in ways consistent with GE comparative statics. Moreover, quantitative GE benchmarks accurately predict final prices and allocations in our markets.

Third, volatility in individual preferences has a predictable influence over the ability of market dynamics to guide prices to equilibrium states. Especially in economies with strong wealth effects, fluctuations in individual preferences can be large relative to excess demands, halting (or even reversing) convergence towards predicted equilibria. By measuring inconsistencies in individual choices in our revealed preference task, we can pre-classify economies as “robust” (likely to converge)

or “fragile” (vulnerable to these fluctuations) before ever observing subjects in markets. These classifications clearly help predict convergence in our markets. What’s more, because we can deliberately construct economies with strong or weak wealth effects simply by sorting subjects into markets based on the convexity of their revealed preferences, we can experimentally control the likelihood of robustness and thereby the reliability of equilibrium convergence.²⁸

Our experiment illustrates some of the ways that studying economies built using “homegrown” preferences can enrich our understanding of competitive behavior, and they highlight connections between markets and revealed preferences that are inaccessible using the usual sort of induced preference procedures. To follow up, we see four important new empirical directions.

First, by design, our experiment precludes the study of consistency violations like GARP (the focus of the extant experimental revealed preferences literature) because we elicit subjects’ preferences with a fixed endowment, preventing budget lines from crossing. Designing experiments that study the relationship between consistency, rationality and aggregate general equilibrium outcomes seems a natural followup direction for this line of work. More generally, studying how systematic errors, bounded rationality and idiosyncratic features of preferences aggregate to markets is an important enterprise for understanding the economic implications of behaviors typically studied at a small, disaggregated scale in the lab.

Second, our sorting algorithm produced 16 non-pathological economies with unique compet-

²⁸An important dimension of our experimental design (to our knowledge new to the literature) is that we elicit subjects’ preferences choices twice on each budget line and show that, on the whole, subjects make remarkably consistent choices (as we show in Online Appendix A.6). This consistency across elicitation lends support to a long-standing conjecture that subjects use decision procedures or rules to organize their choices in settings like this one – it is difficult to imagine that such consistency over so many decisions would be possible for subjects otherwise. The fact that subjects make broadly similar choices also in interactive markets, suggests that such rules may not be just artifacts of elicitation but may also be relevant for guiding decision making in richer contexts.

itive equilibria under which to observe market outcomes (a consequence of the protective effect of diversity on equilibrium uniqueness). This outcome serves our purposes well because it allows us to more precisely test the predictive validity of general equilibrium (it would be more difficult to assess GE predictions in e.g. economies with multiple or nonexistent equilibria). Nonetheless, future experiments designed specifically to produce pathological economies using natural preferences seems valuable for future work. This might be accomplished using richer sorting mechanisms that select relatively homogenous subjects who nonetheless have extremely convex or non-monotonic preferences.

Third, we have so far studied a particular class of preferences, those for risk. Do our largely positive results on aggregation and convergence carry over to other sorts of preferences measurable in the lab (e.g. time preferences)? We had expected homegrown risk preferences to provide a serious challenge to GE, given the strong convexities and frequent violations of neoclassical choice behavior seen in prior experimental risk-elicitation research. Thus we conjecture similarly positive results for GE economies constructed using other sorts preferences, but new empirical surprises surely are possible.

Finally, our experiment was conducted using a particular market institution with features particularly well-suited to the questions motivating our research. The tatonnement institution is implemented using an interface very similar to our revealed preference task, and it is initially adapted to a pre-specified price grid matching the grid used in the Phase 1 elicitation. Perhaps most importantly, tatonnement allows trade only once, after prices have equilibrated; allocations remain constant while the market is open, allowing us to use elicited excess demands to make sharp predictions about dynamics without imposing parametric assumptions on preferences. Studying other, more naturalistic institutions such as the continuous double auction (in which allocations change

during the the trading process) may require more elicitation or stronger structural assumptions, but by the same token may offer deeper insights into market dynamics.

References

- AHN, D., S. CHOI, D. GALE AND S. KARIV, “Estimating ambiguity aversion in a portfolio choice experiment,” *Quantitative Economics* 5 (2014), 195–223.
- ANDERSON, C. M., C. R. PLOTT, K.-I. SHIMOMURA AND S. GRANAT, “Global Instability in Experimental General Equilibrium: The Scarf Example,” *Journal of Economic Theory* 115 (2004), 209–249.
- ANDREONI, J. AND J. MILLER, “Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism,” *Econometrica* 70 (2002), 737–753.
- ASPAROUHOVA, E., P. BOSSAERTS AND C. PLOTT, “Excess Demand And Equilibration In Multi-Security Financial Markets: The Empirical Evidence,” *Journal of Financial Markets* 6 (2003), 1–21.
- ASPAROUHOVA, E., P. BOSSAERTS, N. ROY AND W. ZAME, ““Lucas” in the laboratory,” *Journal of Finance* 71 (2016), 2727–2780.
- BECKER, G. S., “Irrational behavior and economic theory,” *Journal of Political Economy* 70 (1962), 1–13.
- BIAIS, B., T. MARIOTTI, S. MOINAS AND S. POUGET, “Asset pricing and risk-sharing in a complete market: An experimental investigation,” mimeo, 2017.
- BOSSAERTS, P., P. GHIRARDATO, S. GUARNASCHILLI AND W. R. ZAME, “Ambiguity in asset markets: Theory and experiment,” *Review of Financial Studies* 23 (2010), 1325–1359.
- BOSSAERTS, P. AND C. PLOTT, “Basic Principles of Asset Pricing Theory: Evidence From Large-Scale Experimental Financial Markets,” *Review of Finance* 8 (2004), 135–169.

- BOSSAERTS, P., C. PLOTT AND W. ZAME, “Prices and Allocations in Financial Markets: Theory, Econometrics, and Experiments,” *Econometrica* 75 (2007), 993–1038.
- BREABAN, A. AND C. N. NOUSSAIR, “Trader Characteristics and Fundamental Value Trajectories in an Asset Market Experiment,” *Journal of Behavioural and Experimental Finance* 8 (2015), 1–17.
- BRONFMAN, C., K. MCCABE, D. PORTER, S. RASSENTI AND V. SMITH, “An experimental examination of the Walrasian tâtonnement mechanism,” *The RAND Journal of Economics* 27 (1996), 681–699.
- BUDDISH, E. AND J. B. KESSLER, “Bringing real market participants’ real preferences into the lab: An experiment that changed the course allocation mechanism at Wharton,” mimeo, 2018.
- CHAMBERLIN, E. H., “An Experimental Imperfect Market,” *Journal of Political Economy* 56 (1948), 95–108.
- CHOI, S., R. FISMAN, D. GALE AND S. KARIV, “Consistency and Heterogeneity of Individual Behavior under Uncertainty,” *The American Economic Review* 97 (2007), 1921–1938.
- CROCKETT, S., J. DUFFY AND Y. IZHAKIAN, “An Experimental Test of the Lucas Asset Pricing Model,” *Review of Economic Studies* 86 (2019), 627–667.
- CROCKETT, S., R. OPREA AND C. R. PLOTT, “Extreme Walrasian Dynamics: The Gale Example in the Lab,” *The American Economic Review* 101 (2011), 3196–3220.
- DEBREU, G., “Excess Demand Functions,” *Journal of Mathematical Economics* 1 (1974), 15–21.
- DEELSTRA, G., J. DHAENE AND M. VANMAELE, “An overview of comonotonicity and its applications in finance and insurance,” in G. D. Nunno and B. Øksendal, eds., *Advanced Mathematical Methods for Finance* (Springer, Berlin, Heidelberg, 2011), 155–179.

- FISMAN, R., S. KARIV AND D. MARKOVITS, “Individual Preferences for Giving,” *The American Economic Review* 97 (2007), 1858–1876.
- GALE, D., “A Note on Global Instability of Competitive Equilibrium,” *Naval Research Logistics Quarterly* 10 (1963), 81–87.
- GILLEN, B. J., M. HIROTA, M. HSU, C. R. PLOTT AND B. W. ROGERS, “Divergence and Convergence in Scarf Cycle Environments: Experiments and Predictability in the Dynamics of General Equilibrium Systems,” *Economic Theory* (2020), forthcoming.
- GIRAUD, G. AND J. K.-H. QUAH, “Homothetic or Cobb-Douglas Behavior Through Aggregation,” *The B.E. Journal of Theoretical Economics* 3 (2003).
- GJERSTAD, S., “Price Dynamics in an Exchange Economy,” *Economic Theory* 52 (2013), 461–500.
- GOEREE, J. AND L. LINDSAY, “Market Design and the Stability of General Equilibrium,” *Journal of Economic Theory* 165 (2016), 37–68.
- GRANDMONT, J.-M., “Distributions of preferences and the “Law of Demand”,” *Econometrica* 55 (1987), 155–161.
- , “Transformation of the commodity space, behavioral heterogeneity, and the aggregation problem,” *Journal of Economic Theory* 57 (1992), 1–35.
- GUL, F., “A Theory of Disappointment Aversion,” *Econometrica* 59 (1991), 667–686.
- GULER, B., V. LUGOVSKYY, D. PUZZELLO AND S. TUCKER, “Trading institutions in experimental asset markets: Theory and evidence,” mimeo, 2018.
- HALEVY, Y., D. PERSITZ AND L. ZRILL, “Parametric Recoverability of Preferences,” *Journal of Political Economy* 126 (2018), 1558–1593.

- HAMMOND, P. AND S. TRAUB, "A Three-Stage Experimental Test of Revealed Preference," Technical Report 72, Competitive Advantage in the Global Economy (CAGE) Working Paper Series, 2012.
- HARRISON, G., J. LIST AND TOWE, "Naturally occurring preferences and exogenous laboratory experiments; A case study of risk aversion," *Econometrica* 75 (2007), 433–458.
- HILDENBRAND, W., "On the Law of Demand," *Econometrica* 51 (1983), 997–1019.
- , *Market Demand: Theory and empirical evidence* (Princeton University Press, 1994).
- HILDENBRAND, W. AND A. KNEIP, "On behavioral heterogeneity," *Economic Theory* 25 (2005), 155–169.
- JOYCE, P., "The Walrasian *tâtonnement* mechanism and information," *The RAND Journal of Economics* 15 (1984), 416–425.
- LEVY, H., "Risk and return: An experimental analysis," *International Economic Review* 38 (1997), 119–149.
- LIST, J., "Does market experience eliminate market anomalies," *Quarterly Journal of Economics* 118 (2003), 41–71.
- , "The behavioralist meets the market: Measuring social preferences and reputation effects in actual transactions," *Journal of Political Economy* 114 (2006), 1–37.
- MANTEL, R. R., "On the characterization of aggregate excess demand," *Journal of Economic Theory* 7 (1974), 348–353.
- MAS-COLELL, A., "On the equilibrium price set of an exchange economy," *Journal of Mathematical Economics* 4 (1977), 117–126.

- PLOTT, C. R., “Market Stability: Backward-bending supply in a laboratory experimental market,” *Economic Enquiry* 38 (2000), 1–18.
- PLOTT, C. R. AND G. GEORGE, “Marshallian vs. Walrasian stability in an experimental market,” *The Economic Journal* 102 (1992), 437–460.
- SCARF, H., “Some examples of global instability of the competitive equilibrium,” *International Economic Review* 1 (1960), 157–172.
- SHAPLEY, L. S. AND M. SHUBIK, “An Example of a Trading Economy with Three Competitive Equilibria,” *Journal of Political Economy* 85 (1977), 873–875.
- SMITH, V. L., “An experimental study of competitive market behavior,” *Journal of Political Economy* 70 (1962), 111–137.
- SMITH, V. L., G. L. SUCHANEK AND A. W. WILLIAMS, “Bubbles, crashes, and endogenous expectations in experimental spot asset market,” *Econometrica* 56 (1988), 1119–1151.
- SONNENSCHN, H., “Do Walras’ identity and continuity characterize the class of community excess demand functions?,” *Journal of Economic Theory* 6 (1973), 345–353.
- WALRAS, M.-E.-L., *Elements of Pure Economics* (Translated, Harvard University Press (1954), 1877).
- WILLIAMS, A., V. L. SMITH, J. LEDYARD AND S. GJERSTAD, “Concurrent trading in two experimental markets with demand interdependence,” *Economic Theory* 16 (2000), 511–528.