It's complicated: A Non–parametric Test of Preference Stability between Singles and Couples

Stefan Hubner[1]

[1] *Nuffield College and Department of Economics, University of Oxford**

September 20, 2018

Abstract *This paper develops a non–parametric test of the controversial preference stability assumption. This assumption is often made in order to use data from single households as an identification strategy in the collective model. Our test allows for unobserved heterogeneity by defining finite-dimensional types of households according to their revealed preference relations. We show how to derive a test statistic by constructing hypothetical matches of heterogeneous individuals into different types of households using tools from stochastic choice theory. We strongly reject the preference–stability hypothesis based on data from the Russian Longitudinal Monitoring Survey (RLMS) and the Spanish Continuous Family Expenditure Survey (ECPF).*

JEL Codes: *D12, D13, J12*
Keywords: *Collective model, Preference Stability, Collective Axiom of Revealed Preference, Stochastic Choice, Random Utility Models*

1 Introduction

The classical household consumption model, which assumes that a household consists of only one decision maker, is unable to answer many policy-

*Stefan Hubner (stefan.hubner@economics.ox.ac.uk) gratefully acknowledges financial support from the Netherlands Organisation for Scientific Research (NWO) through a TOP Grant. The author would also like to thank Arthur van Soest, Laurens Cherchye, Frederic Vermeulen, Stefan Hoderlein, Yuichi Kitamura, Jörg Stoye, John Quah, Marting Browning and various participants from the Heterogeneity in Supply and Demand workshop (2016) for their helpful comments.

relevant questions concerning family economics, such as pooling of taxable income or deciding which partner should receive the payments of childcare benefits. The collective household consumption model (Chiappori, 1988, 1992) provides an alternative to the unitary model by allowing household members to efficiently bargain over their consumption choices. However, even if individuals have homogeneous preferences, the proportion at which households split their resources is not identified from household–level data (Chiappori & Ekeland, 2006, 2009). Researchers often circumvent this issue by making use of data from singles and assume that the transition from living in a single household to cohabitation as a couple and vice versa does not affect individual's consumption preferences (most prominently Browning, Chiappori & Lewbel (2013), but also Barmby & Smith (2001), Vermeulen (2006), Bargain & Donni (2012), Gayle & Shephard (2016)).

In this paper, we construct a test of the validity of this *stable preference* assumption. In order to avoid rejecting the hypothesis based on restrictions imposed by the functional form of utilities or by preference homogeneity, our test is fully non–parametric and allows for a heterogeneous population. Preference homogeneity is particularly restrictive in the context of the collective model since it not only requires every individual to have the same preferences but also assumes that any two individuals matched as a couple would arrive at the exact same sharing of resources. Here, we specify a collective random utility model with continuous consumption and an arbitrary dimension of unobserved distribution factors[1] and preference parameters. We then construct discrete heterogeneous household types for both couples and singles in a way that ensures that any two households which are not distinguishable in terms of their preferences without a functional form restriction are equivalent. The test is then constructed by considering couples satisfying CARP, the Collective Axiom of Revealed Preference (Cherchye, De Rock & Vermeulen, 2007), and singles satisfying SARP, the Strong Axiom of Revealed Preference (Afriat, 1967; Varian, 1982), as a baseline. Observing only their respective distributions of

[1]Unobserved distribution factors are random variables effecting the bargaining power, but not the preferences of an individual; for example desirability on the marriage market (Hubner, 2018).

consumption choices, we ask the question if these distributions could have been generated from an unobserved combination of hypothetical matches between different types of individuals into a common household. This requires the stable preference assumption so that failing to find a stochastic rationalisation of the data leads to a rejection of the stable preference hypothesis.

A difficulty when using or testing preference stability is the presence of consumption externalities. Without further assumptions, they prohibit disentangling adjustments of consumption behaviour due to preference changes from the acquired possibility of consuming a good publicly as a couple, i.e. changes in Lindahl prices. For example, consider individuals with stable preferences commuting to work by car. While as a single they have to pay market prices for gasoline, as a couple they can share the burden and consequently consume more other goods, which could lead us to believe that preferences have changed. Thus, we consider a generalisation of a Beckerian caring model (Becker, 1981), which restricts us to only consider strictly private goods for the empirical test. In order to identify preference relations for singles and couples, we make use of panel data and a common time-homogeneity of preferences assumption. We apply our test to two popular datasets: the *Russian Longitudinal Monitoring Survey (RLMS)* and the *Spanish Continuous Family Expenditure Survey (Encuesta Continua de Presupuestos Familiares, ECPF)* which were used by Cherchye, De Rock & Vermeulen (2011) and Adams *et al.* (2014) in the context of the collective model. We consistently reject the hypothesis of preference stability for both datasets.

In contrast to the discrete approach, we propose in this paper, testing preference restrictions in a continuous setting is often based on the Slutsky matrix which requires estimation of household demands and their derivatives. Browning & Chiappori (1998) construct a test of collective rationality based on a parametric almost ideal demand system with additive measurement errors. Similar to this, Brugler (2016) estimates a parametric quadratic ideal demand system (Banks, Blundell & Lewbel, 1997) in a setting without preference heterogeneity and compares the parameter estimates for single men, single women and couples to draw conclusions about preference stability. While almost ideal demand systems provide a flexible functional parametric form where parameter restrictions can be easily tested, the potential for model misspecification and consequential type I errors caused by an inconsistent specification of the functional form restriction can be problematic. In addition to this, even in a fully parametric setting, identification of continuous demand systems in the presence of general unobserved heterogeneity is difficult due to the structure of the collective model, which leads to demands that are non–separable with respect to unobserved preference and bargaining heterogeneity. Also in a continuous choice setting, Hubner (2018) develops a non–parametric collective random utility model and derives restrictions for non–parametric identification of idiosyncratic utility functions and Pareto weights by showing global invertibility of demands, under the assumption of observed private demands. We do not require such a strong form of identification for our test.

To our knowledge, the use of singles data in the context of the fully non–parametric way to model collective households using revealed preference restrictions (Cherchye, De Rock & Vermeulen, 2007, 2009) is novel. The advantage of a revealed preference based approach is that it allows us to use a stochastic random utility and random distribution factor version of the collective model without requiring global invertibility. Revealed stochastic preference settings have recently been used in the context of the unitary consumption model. Hoderlein & Stoye (2014) consider the weak axiom of revealed preference in the unitary model. In particular, they use the fact that demands of a heterogeneous population observed in a given price regime can be characterised as random variables supported on a normalised budget set. Observing the same population in different price regimes (repeated cross-sections), one can then use copula techniques to derive (Frechet–Hoeffding) bounds on the probability that the population behaves irrationally, i.e. is not in line with the weak axiom. Kitamura & Stoye (2013); Deb *et al.* (2017) integrate this approach into the stochastic choice framework of McFadden & Richter (1991) and McFadden (2005) by fully discretising budget sets using partitions which contain all the relevant information to test the strong axiom of revealed preference.

The construction of our test statistic is closely related to these approaches. We show that the large sample theory developed in Kitamura & Stoye (2013) applies to our test statistic and use their results for our statistical inference. We complement this by implementing a fast, parallel non-negative least squares algorithm which leverages the sparsity of the computational problem in Haskell and conduct a simulation study to evaluate finite sample size and power of the test statistic in our setting.

2 Test Design

We start by specifying a collective random utility model with two-person households. Each spouse $r \in \{f, m\}$ consumes a bundle of goods from a finite set of alternatives which is a proper subset of $\mathbb{R}^L_+$. We denote continuous individual private consumption by x^r. Further, let $x^c_{i,t} = x^f_{i,t} + x^m_{i,t} \in R^L_+$ be continuous household consumption chosen by household $i \in I_N$ in period $t \in I_T$. We assume that a household, characterised by observed (p_t, w_t) and unobserved $\varepsilon^c = (\varepsilon^m, \varepsilon^f, \varepsilon^\mu)$, arrives at this consumption bundle by having maximised its collective random utility

$$\max_{x^f, x^m} \left\{u^m(x^m, x^f, \varepsilon^m) + \mu(p, \varepsilon^\mu)u^f(x^f, x^m, \varepsilon^f)\right\} \text{ subject to } x^f + x^m \in B_t = \{x \mid p_t x \leqslant w_t\}, \tag{1}$$

where μ is the relative bargaining power of spouse f (Chiappori, 1988, 1992). By introducing the possibly infinite dimensional random variable ε^c we allow each household to optimise a different objective function according to idiosyncratic preferences and distribution factors. For each period $t \in I_T$ we observe expenditures for a given good $l \in I_L$ denoted by $p_{t,l}x^c_{i,t,l}$. Observing prices p_t then allows us to calculate the vector of continuous household consumption $x^c_{i,t}$ for each household $i \in I_N$.

Now consider instead a single household $r \in \{f, m\}$ who, under the stable preference assumption, maximises the same $u^r(x^r, x^{r'}, \varepsilon^r)$ subject to the constraint $x^r \in B_t$ according to the standard unitary model. For them, the spouse's consumption $x^{r'}$ is zero. Thus for single households we conveniently observe $x^c = x^r$. Conversely, due to potential complementaries arising from joint consumption of a good with a potential partner, without further restrictions, these zeros will cause an ill-posed utility maximisation problem for single households. Thus, in order to model singles in a way that makes them informative for a couple's consumption behaviour one has to make the following separability assumption.

Assumption 1. *Let the* $(L-1)$*-dimensional vector of marginal rates of substitutions for* $r \in \{f, m\}$ *be denoted as* $MRS^r(x^r, x^{r'})$ *with components* $MRS^r_l(x^r, x^{r'}) = \frac{\partial u^r/\partial x^r_l}{\partial u^r/\partial x^r_L}$ *for* $l = 1, \ldots, L-1$. *Then for* $r \neq r'$ *we have* $\partial MRS^r(x^r, x^{r'})/\partial x^{r'} = 0_{L-1,L-1}$.

This assumption is standard and required whenever singles are used as an identification strategy for individual preferences in the collective model. It states that the marginal rates of substitution for own good consumption does not depend on the spouse's consumption. A sufficient condition for this is separability of the form $u^r(x^r, x^{r'}, \varepsilon^r, \varepsilon^{r'}) = G(g(x^r, \varepsilon^r), x^{r'}, \varepsilon^{r'})$ for any two differentiable, increasing, real-valued functions G and g. While this assumption allows for positive consumption externalities, it restricts the way the behaviour of a person is altered when entering or exiting a relationship. For example, it rules out non-cooperative, strategic behaviour of individuals towards their spouse. This assumption nests popular specifications such as the egoistic model $u^r = g^r$, but also the Beckerian caring model with altruistic preferences (Becker, 1981). In this specification, utilities of one spouse are defined in terms of own-good consumption and the utility of the spouse, i.e. $u^r(x^r, \varepsilon^r) = W(U^r(x^r, \varepsilon^r), U^{r'}(x^{r'}, \varepsilon^{r'}))$ where U^f and U^m are real-valued subutility functions with the usual properties and W is a strictly increasing, differentiable real-valued function.[2] We make this assumption to disentangle separability from stable preferences, or otherwise, we would test both assumptions jointly. In either case, a rejection implies that singles should not be used for identification of the sharing rule.

In the collective model, demands of each spouse are in general not observ-

[2]We can write $u^r(x^r, \varepsilon^r) = u^r(x^r, x^{r'}, \varepsilon^r, \varepsilon^{r'})$ because the Beckerian model is observationally equivalent with egoistic individual utility functions in the collective model.

able and thus their individual preferences are not identified directly from data of cohabiting couples. Using the *stable preference assumption*, how can we exploit information about single households in order to identify u^f and u^m?

To answer this question, we first characterise consumers in terms of their finite-dimensional revealed preference relations. Both the collective and the unitary model impose restrictions on how both household and individual demands must change with respect to relative price changes. These sets of restrictions are known as the Collective Axiom of Revealed Preference and the Strong Axiom of Revealed Preference, both defined in Appendix A.1. The axioms allow us to partition unobserved heterogeneity in the following way. Let $\mathcal{E}^r = \bigcup_{k_r=1}^{K_r} \mathcal{E}_{k_r}$ such that for all $r \in \{c, f, m\}$ and for all $k_r \in I_{K_r}$ it holds that for any $\varepsilon^r_{k_r}, \xi^r_{k_r} \in \mathcal{E}_{k_r}$ we have $R^r(\varepsilon^r_{k_r}) = R^r(\xi^r_{k_r})$ where R^r are the preference relations resulting from our random utility model (1) with household utility evaluated at $u^m(x^m, \varepsilon^m_{k_m}) + \mu(p, \varepsilon^c_{k_c})u^f(x^f, \varepsilon^f_{k_f})$. Consequently, without losing any important information, we map infinite dimensional unobserved heterogeneity into a finite-dimensional collection of graphs representing all possible reference relations R^f, R^m and R^c.[3] We can directly identify them from data by making the following standard assumption (Cherchye, De Rock & Vermeulen, 2007, 2011):

Assumption 2. *(i) Unobserved preferences and distribution factors are constant over time, so that* $\varepsilon^r_i = \varepsilon^r_{i,t}$ *for all* $r \in \{c, f, m\}$, $t \in I_T$ *and* $i \in I_N$.

(ii) We observe choices for each household (couples and singles) for at least three periods.

The time-homogeneity assumption of preferences is needed so that we can treat different periods as different price regimes. To be more precise, we have to assume that preferences do not change over time, such that we can treat the heterogeneity of choices between periods $t \in I_T$ as a consequence of individuals facing different prices p_t, rather than a change in preferences over

[3] We use $R^c = R^c_0$ as notation. It does not actually represent a preference relation, since household consumption is only a result of individual preferences.

time.[4,5]

Let $R^f, R^m \in \mathcal{X}^m = \mathcal{X}^f$ and $R^c \in \mathcal{X}^c$. While, the construction of the type space $\mathcal{X}^f$ is straightforward, the constructing $\mathcal{X}^c$ requires an extension of the space of goods.[6] Considering a finite number of choice types allows us to fully characterise hypothetically matched households, by considering the product space $\mathfrak{X} = \mathcal{X}^c \times \mathcal{X}^f \times \mathcal{X}^m$. This should be interpreted as matching different consumption types $R^f(\varepsilon^f)$ and $R^m(\varepsilon^m)$ into different types of bargaining agreements $R^c(\varepsilon^c)$. Under the preference stability assumption, which ensures that single male and single female households are informative for the respective spouse's behaviour within a couple, each match, characterised by the three-tuple (R^c, R^m, R^f), should behave rationally according to the collective model. Figure 1 presents a clustered graph representation of the type space for matched households. The edges between the nodes represent hypothetical matches and are *not* observed from data.

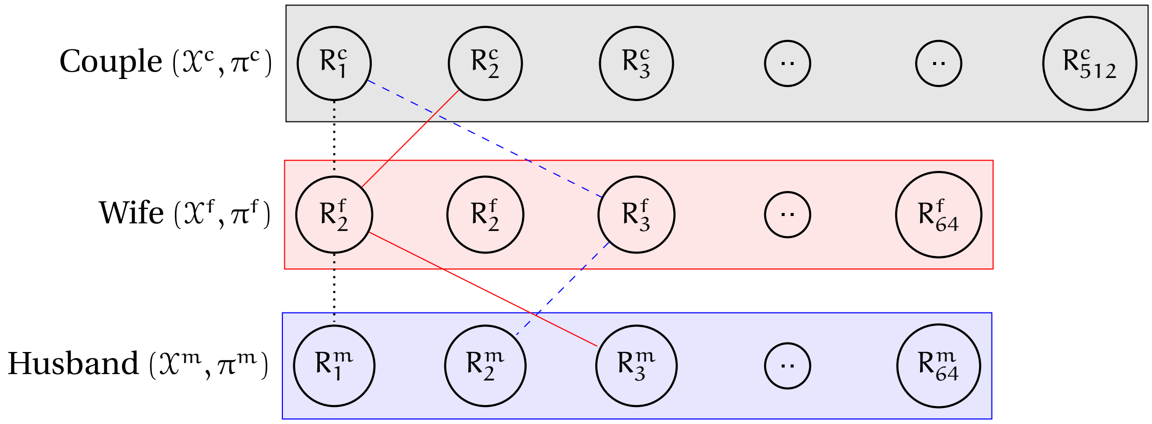
Thus, in a second step, we show that by using information about singles we can fully validate the revealed preference axioms for a heterogeneous population while only observing marginal distributions of choices for households with different compositions. To see this, let the probability that option $\xi_j \in \mathcal{X}^r$ is chosen by a household of a given composition r be denoted as $\pi(\xi_j|\mathcal{X}^r)$. By definition of our type space, this is equivalent to the probability of a household being of type R^r_j. We call this a stochastic choice in situation $\mathcal{X}^r$ and will often refer to it as the marginal distribution of choices under a given household composition. We can consistently estimate these distributions from sample frequencies. Using principles from stochastic choice theory (McFadden & Richter, 1991; McFadden, 2005), we can ask the question whether there ex-

[4]This assumption can likely be relaxed to continuous demands of the form $x_{it} = x(p_t, \varepsilon_i) + \varepsilon_{it}$, where ε_i is an element of a general probability space capturing unobserved heterogeneity and ε_{it} is a period specific idiosyncratic taste shock. Identification of such a specification is treated in Evdokimov (2010).

[5]Identification of the preference relation from multiple periods also has the advantage that we do not have to project observed demands onto budget sets to reduce the dimension of the choice space, such as Kitamura & Stoye (2013), which theoretically requires a homothetic preference assumption or a revealed price preference setting; c.f. Deb *et al.* (2017).

[6]This is due to the requirement of embedding information about double sums to evaluate CARP restrictions (iv) & (v) in Definition 1 of Appendix A.1. A discussion of the construction of type spaces for a minimal economy can be found in Appendix A.3.

Figure 1: Visualisation: Stochastic Collective Revealed Preferences



Note: 3-partite graph where the nodes represent discrete consumption decisions (revealed preference types) and the partitioning is such that there are three disjoint classes each representing the set of discrete decisions under a given household composition – single female f, single male m and couples and c. The marginal distributions π^c, π^f and π^m, supported on $\mathcal{X}^c$, $\mathcal{X}^f$ and $\mathcal{X}^m$, respectively, are identified from data. Edges between the type represent hypothetically matched households. Every possible match, represented by a tuple of edges, can be classified as rational or irrational according to a given set of restrictions.

ists a probability measure ν over an appropriate subset of matched household types $\mathfrak{X}^0 \subset \mathfrak{X}$ that rationalises the observed stochastic choices $\pi(\xi_j|\mathcal{X}^r)$ for $r \in \{c, f, m\}$. With this construction, the choice function $\mathcal{X} \mapsto \Xi(\mathcal{X})$, determining a decision rule for a given state of the world, is a function of household composition $\mathcal{X} \in \{\mathcal{X}^c, \mathcal{X}^f, \mathcal{X}^m\}$. We say that the stochastic choice π is stochastically rational if, for all household compositions $\mathcal{X}^r$ for $r \in \{c, f, m\}$, it holds that $\pi(\xi_j|\mathcal{X}^r) = \nu(\{\Xi \in \mathfrak{X}^0 : \xi_j = \Xi(\mathcal{X}^r)\}$. Intuitively, if the choices in different states of the world (hypothetical matches) can be rationalised by a probability distribution over a set of rational matched household types (edges consistent with Ξ), we can say that the population is rational with respect to the decision rule Ξ.

Finally, how can we test the preference stability assumption by choosing the appropriate decision rule? For this, we partition the universe of types $\mathfrak{X} = \mathfrak{X}^{\text{collective}} \cup \mathfrak{X}^{\text{alternative}}$, where the set $\mathfrak{X}^{\text{collective}}$ contains all types of couples R^c for which there exists a tuple (H^f, H^m), which is consistent with the collective axiom as in Cherchye, De Rock & Vermeulen (2007, 2011). In order to use the actual respective preference relations (R^f, R^m) for a given matched household type, we require the preference stability assumption. We denote the subset of households who remain consistent with the collective axiom after replacing (H^f, H^m) by (R^f, R^m) as $\mathfrak{X}^0 = \mathfrak{X}^{\text{collective}} \setminus \mathfrak{X}^1$. Its complement $\mathfrak{X}^1$ then consists of the cases which satisfy the collective axiom based on couples data but are not consistent with the collective axiom if preference relations from singles are added. Thus, for our test, we drop all cases that are not collectively rational and test $\mathfrak{X}^0$ against $\mathfrak{X}^1$ to answer the question whether the stable preference assumption holds.[7] If we find that among the collectively rational couples it is not possible to rationalise observed choice probabilities using the types belonging to the set $\mathfrak{X}^0$, then we can conclude that the hypothesis of stable preferences does not hold.

3 Test Statistic

To construct a test statistic, it will be useful to represent the abstract choice rule Ξ, as something more traceable. McFadden & Richter (1991); McFadden (2005) show that stochastic choices can be represented by a linear system of equations. Lemma 1 summarises and extends some of these results.

Lemma 1. *The following statements are equivalent:*

(i) The population defined by choice distribution π and choice rules belonging to $\mathfrak{X}^0$ according to the random collective utility model (1) *is rational and satisfies the stable preference assumption.*

(ii) There exists $\nu \in \Delta^{|\mathfrak{X}^0|}$ such that $A\nu = \pi$ where Δ^M is the M-dimensional probability simplex and where the columns of A represent an exhaustive list of rational types according to (1) *under the preference stability assumption.*

[7] Alternatively, one could test $\mathfrak{X}^{\text{collective}}$ against $\mathfrak{X}^{\text{alternative}}$. In fact, this is what Cherchye, De Rock & Vermeulen (2007) do which does not require single data since such a test can be based on aggregate consumption only and the whole joint distribution of such choices is directly identified from data. Equally, by adding single data and assuming separable and stable preferences one could test $\mathfrak{X}^0$ against $\mathfrak{X}^1 \cup \mathfrak{X}^{\text{alternative}}$ to obtain a stronger test of the collective model compared to the previous one. While this test has more power, it comes with the drawback of only applying to a separable caring-type model.

(iii) *Let* $\underline{\nu} = 0$. *Then the vector* ν *solves* $\mathcal{J}_N(\pi, \underline{\nu}) := N \min_{\eta \in \{A\nu | \nu \geqslant \underline{\nu}\}} (\pi - \eta)^T \Omega (\pi - \eta) = 0$.

(iv) *Similarly, for* $\underline{\nu} = 0$ *the vector* ν *is a fixed point under the operation*

$$\Psi_{\pi,\underline{\nu}} : s \mapsto \mathit{max}(0, s - \mathit{diag}(H\iota)^{-1}(H^T s + f(\pi, \underline{\nu})) \tag{2}$$

where $H = A^T \Omega A$ *and* $f(\pi, \underline{\nu}) = -A^T \Omega (\pi - A\underline{\nu})$.

Proof. See Appendix A.2 □

In order to construct the matrix A, we consider a matrix with $\sum_{r \in \{c,f,m\}} |\mathcal{X}^r|$ rows and $|\mathfrak{X}^0|$ columns, where $|\mathcal{X}^r|$ is the number of choices a household can make under a given composition. We then split all columns $A_{\bullet,k}$ with $k \in I_{|\mathfrak{X}^0|}$ into 3 blocks of respective length $|\mathcal{X}^c|$, $|\mathcal{X}^f|$ and $|\mathcal{X}^m|$ and denote each block by $A_{r,\bullet,k}$. If household match $k \in I_{|\mathfrak{X}^0|}$ is of type j under composition $r \in \{c, f, m\}$ then $A_{r,j,k} = 1$ and zero otherwise. In the graph interpretation of the type space in Figure 1, a block refers to a cluster of the graph and a match is represented by a column of the matrix A, where a row value of 1 indicates the active node.

We obtain choice probabilities by consistently estimating the empirical distribution of finite-dimensional household types from the continuous distribution of consumption for each household $i \in I_N$ and $r \in \{c, f, m\}$. Let π be a vector with $\sum_{r \in \{c,f,m\}} |\mathcal{X}^r|$ rows representing observed choice probabilities. Partitioning π the same way as a column A_k, for $r \in \{c, f, m\}$ we define $\pi_{r,j} = \frac{1}{N_r} \sum_{i \in I_{N_r}} \sum_{R^r \in \mathcal{X}^r} \mathbb{1}\{R_i^r = R^r\}$ where R_i^r is the type (preference relation) of household $i \in I_{N_r}$ and our sample is partitioned as $I_N = I_{N_c} \cup I_{N_f} \cup I_{N_m}$.

The way the matrix A is constructed, ν is not point-identified in $A\nu = \pi$ since A is far from full column rank with $|\mathfrak{X}^0| \gg \sum_{r \in \{c,f,m\}} |\mathcal{X}^r|$. Thus we make use of the equivalence between (ii) and (iii) in Lemma 1 and estimate η by projecting observed choice probabilities $\hat{\pi}$ onto a τ_N-tightened[8] linear cone $\mathcal{C} = \{A\nu : \nu \geqslant \tau_N \iota\}$ by minimising the projection residuals $\mathcal{J}_N(\hat{\pi}, \iota\tau_N)$. We set $\tau_N = \frac{1}{H}\sqrt{\frac{\log \underline{N}}{\underline{N}}}$ and $\underline{N} = N_f \wedge N_m \wedge N_c$ is the minimum number of available observations per household composition N_r for $r \in \{c, f, m\}$ in the sample. To

[8]Tightening is required since many of the inequality constraints describing the cone will be binding. With the parameter being on the boundary of the parameter space the bootstrap procedure we use would not be valid; Andrews (2000).

obtain the critical value, we then use a non–parametric bootstrap to obtain $\widehat{\pi}^b$ and calculate $\mathcal{J}_N\left(\widehat{\pi}^b, \iota\tau_N\right)$ for each $b \in I_B$, where B the number of bootstrap repetitions. Let $\widehat{\eta}_{\tau_N}$ be the argument producing the projection residuals $\mathcal{J}_N(\widehat{\pi}, \iota\tau_N)$. The centred choice probabilities $\widehat{\pi}^b_{\tau_N} = \widehat{\pi}^b - \widehat{\pi} - \widehat{\eta}_{\tau_N}$ are then used to approximate the empirical distribution $F_{\mathcal{J}_N}$ of $\mathcal{J}_N(\widehat{\pi}, \iota\tau_N)$. Then, following Kitamura & Stoye (2013), the bootstrap is valid and we have for $\alpha \in (0, \frac{1}{2})$ and $\tau_N\sqrt{N} \to \infty$

$$\lim_{N\to\infty} \inf \inf_{\pi\in\mathcal{C}} \mathbf{P}\left(\mathcal{J}_N(\widehat{\pi}, \tau_N\iota) \leqslant \widehat{F^{-1}_{\mathcal{J}_N}}(1-\alpha)\right) = 1-\alpha. \tag{3}$$

The projection residuals are calculated for each bootstrap repetition and it will prove useful to rewrite Lemma 1.(iii) as the solution of a non-negative least squares problem and implement a fast algorithm for solving it. The most commonly used method to solve this is sequential quadratic programming, c.f. Lawson & Hanson (1995).[9] Due to the high dimensionality of our problem, it is preferable to use coordinate-wise projection such as in Franc, Hlavac & Navara (2005) or Landweber's gradient descent method (Johansson *et al.*, 2006) approach as it requires only $\mathcal{O}(k)$ computations instead of $\mathcal{O}(k^3)$, where $k = |\mathcal{X}^0|$ is the (generally very large) number of rational types in the NNLS problem. Equation (2) in Lemma 1.(iv) represents a step using Landweber's method, which we implement manually, in order to leverage the sparsity of the matrix A.

Appendix A.4 discusses the results of a simulation study in which we evaluate the power of our test by plotting empirical rejection frequencies against the proportion of households not optimising according to a given decision rule and its size by evaluating type 1 errors under worst case scenarios. We find that the test has power to detect an "irrational" population of close to one with 500 observations per household composition if only 15% of the sample do not behave according to the model. If the sample size is doubled, the required proportion drops to 5%. In addition to this, we discuss worst cases by considering "similar matches" and show that the size of the test is correct under different worst-case samples.

[9]This algorithm is used for `lsqnonneg` in Matlab and `optimize.nnls` in SciPy.

4 Results

In this section, we apply our test to two widely used datasets: The Russian Longitudinal Monitoring Survey (RLMS) and the Spanish Continuous Family Expenditure Survey (ECPF). Neither of these datasets have information about allocation of consumption between the spouses. The results are consistent between the datasets. In both cases, we strongly reject the hypothesis of stable preferences.

For the test we consider households consisting of singles or couples. We exclude households with children or other cohabiting groups of individuals who are not in a romantic relationship. We consider a minimal setting with three periods and three goods, where we have 64 types of singles and 512 types of couples, resulting in 2,996 collectively rational matched household types who satisfy the stable preference assumption.[10] Both panels are longer than required. Thus we evaluate different combinations of years and goods from a range of pre–defined private goods. After dropping incomplete and boundary cases, we select years and goods based on the resulting sample size. Due to attrition in panels, this procedure tends to pick out consecutive years. As such we face the trade–off between a small sample size and small price variation, where the latter decreases the power in any revealed preference setting (see e.g. Beatty & Crawford (2011)). Thus, in addition to the combination with the largest sample size, we report a range of such combinations for a fixed group of goods, in descending order of $N_{\text{couples}} + N_{\text{singles}}$.

We first consider phase two of the Russian Longitudinal Monitoring Survey (RLMS), collected in form of a personal interview by the Carolina Population Center (University of North Carolina) and available for the years 1994 – 2014. Due to a lot of missing or zero values of other types of private consumption expenditures, we focus on different categories of food.[11] The survey distinguishes between 57 different food consumption categories, which we further aggregate to dairy, bread and meat. These three categories account for more than half of the food consumption. Price data is obtained from the Federal

[10] See Appendix A.3 for a detailed discussion of this setting.

[11] The median household income is RUB 28,000, of which RUB 7,500 is spent on food.

State Statistics Service (GKS) and available for the years: 2000, 2005, 2010, 2011 – 2015. Descriptive statistics can be found in Table 4 in Appendix A.5. Table 1 shows the result in form of p-values of our test for different combinations of periods.

Table 1: Results RLMS: Preference Stability Test

Years	$N^{total}_{couples}$	$N^{rational}_{couples}$	$N^{total}_{singles}$	$N^{rational}_{singles}$	p-value	
2012 2013 2014	327	315	305	295	0.027	**
2011 2013 2014	316	304	281	275	0.017	**
2011 2012 2014	314	307	283	276	0.020	**
2011 2012 2013	332	321	312	306	0.187	
2010 2013 2014	258	249	217	213	0.033	**
2010 2012 2014	256	248	221	217	0.060	*
2010 2012 2013	268	257	243	235	0.043	**
2010 2011 2013	268	260	239	235	0.057	*
2010 2011 2012	298	293	268	262	0.120	
2005 2011 2012	264	256	220	214	0.180	

Note: Number of total couples, rational couples according to aggregate CARP, total singles and rational singles according to SARP, for a given combination of years. P-values and significance levels 10%, 5%, and 1% indicated by *, **, and ***, respectively.

Second, to validate these results, we apply the test to data from the Spanish Continuous Family Expenditure Survey (ECPF), collected by the Spanish statistics office (INE) on a quarterly basis for the period 1985 – 2005. The survey is designed in a way that participants are part of the sample for at most eight consecutive periods or two years. There was a discontinuity in the design of the study in 1997, where the focus was shifted away from detailed consumption expenditure categories. While the ECPF was replaced by the Encuesta de Presupuestos Familiare (EPF) in 2006, unfortunately the collection frequency of this newer version was extended to once per year while the participation lifespan of two years was maintained. Requiring a panel of at least three periods, we will, therefore, use data from the original ECPF from 1985 to 1996. We use the following private goods: clothing, food consumed outside of the household and consumption of non–durable articles. Price data is also published by INE. Descriptive statistics can be found in Table 5 in Appendix A.5, Table 2 presents the results.

Table 2: Results ECPF: Preference Stability Test

Years	$N^{total}_{couples}$	$N^{rational}_{couples}$	$N^{total}_{singles}$	$N^{rational}_{singles}$	p-value	
19943 19941 19942	111	106	6	6	0.007	***
19934 19941 19942	111	106	8	8	0.000	***
19922 19923 19921	97	88	16	13	0.000	***
19904 19911 19912	98	94	14	13	0.020	**
19893 19891 19892	113	105	6	6	0.000	***
19882 19883 19884	106	103	7	5	0.000	***
19871 19872 19873	131	128	8	6	0.000	***
19864 19871 19872	164	157	9	6	0.003	***
19863 19864 19871	141	135	8	7	0.000	***
19863 19864 19862	135	130	9	8	0.007	***

Note: Number of total couples, rational couples according to aggregate CARP, total singles and rational singles according to SARP, for a given combination of years. P-values and significance levels 10%, 5%, and 1% indicated by *, **, and ***, respectively.

Two aspects of our results are worth noting. First, while there is overwhelming evidence to reject the stable–preference hypothesis for the RLMS (Table 1), there are some combination of periods for which there is not enough evidence to arrive at this conclusion. In any of these cases, we either have three consecutive years in which we are faced with little power of revealed preference axioms due to the lack of price variation, or a particularly small sample size due to the wide span of considered years. This all points towards the trade–off discussed above. Second, Table 2 shows that our sample for the ECPF is very small, particularly for single households. Abstracting from the inferior statistical properties of the test in small samples (the numerical procedure still converges), the strong rejection of the hypothesis seems to indicate that it is harder to find a rationalisation of types when we observe zero probability mass for some types $R \in \mathcal{X}^r$ for some $r \in \{c, f, m\}$. As a standalone result, we would have to take the result with some caution. However, due to the consistency with the result from the first dataset, we believe that the evidence uniformly points towards a rejection of the stable–preference hypothesis.

5 Conclusion

In this paper, we constructed a test for the hypothesis of stable preferences. For this, we set up a collective random utility model and a unitary random utility model and used a discretisation of continuous choices to revealed preference types for both types of households. We then asked the question under which conditions we can construct hypothetical matches of different heterogeneous individuals into different types of households. In a caring model, this is possible under the assumption of stable preferences, which formed the basis for our test. By considering collectively rational couples and a stochastic choice argument, we then showed that under the null hypothesis of stable preferences, there exists a stochastic rationalisation of observed choice data. Using data from the Russian Longitudinal Monitoring Survey (RLMS) and the Spanish Continuous Family Expenditure Survey (ECPF), we strongly rejected the hypothesis.

References

Abi ADAMS, Laurens CHERCHYE, Bram DE ROCK & Ewout VERRIEST (2014), Consume now or later? Time inconsistency, collective choice, and revealed preference. *American Economic Review*, vol. 104, no. 12, pp. 4027–4070.

Sydney AFRIAT (1967), The construction of utility functions from expenditure data. *International Economic Review*, vol. 8, no. 1, pp. 67–77.

Donald ANDREWS (2000), Inconsistency of the Bootstrap when a Parameter is on the Boundary of the Parameter Space. *Econometrica*, vol. 68, no. 2, pp. 399–405.

James BANKS, Richard BLUNDELL & Arthur LEWBEL (1997), Quadratic Engel Curves and Consumer Demand. *Review of Economics and Statistics*, vol. 79, no. 4, pp. 527–539.

Olivier BARGAIN & Olivier DONNI (2012), Expenditure on children : A

Rothbarth-type method consistent with scale economies and parents bargaining. *European Economic Review*, vol. 56, no. 4, pp. 792–813.

Tim BARMBY & Nina SMITH (2001), Household labour supply in Britain and Denmark: some interpretations using a model of Pareto Optimal behaviour. *Applied Economics*, vol. 33, no. 9, pp. 1109–1116.

Timothy BEATTY & Ian CRAWFORD (2011), How demanding is the revealed preference approach to demand? *American Economic Review*, vol. 101, no. 6, pp. 2782–2795.

Gary S. BECKER (1981), Altruism in the Family and Selfishness in the Market Place. *Economica*, vol. 48, no. 189, pp. 1–15.

Martin BROWNING & Pierre-Andre CHIAPPORI (1998), Efficient Intra-Household Allocations: A General Characterization and Empirical Tests. *Econometrica*, vol. 66, no. 6, pp. 1241–1278.

Martin BROWNING, Pierre-Andre CHIAPPORI & Arthur LEWBEL (2013), Estimating consumption economies of scale, adult equivalence scales, and household bargaining power. *Review of Economic Studies*, vol. 80, no. 4, pp. 1267–1303.

James BRUGLER (2016), Testing preference stability between couples and singles. *Economics Letters*, vol. 142, pp. 15–17.

Laurens CHERCHYE, Bram DE ROCK & Frederic VERMEULEN (2007), The collective model of household consumption: a nonparametric characterization. *Econometrica*, vol. 75, no. 2, pp. 553–574.

Laurens CHERCHYE, Bram DE ROCK & Frederic VERMEULEN (2009), Opening the black box of intra-household decision-making. *Journal of Political Economy*, vol. 117, no. 6, pp. 1074–1104.

Laurens CHERCHYE, Bram DE ROCK & Frederic VERMEULEN (2011), The Revealed Preference Approach to Collective Consumption Behaviour: Testing and Sharing Rule Recovery. *The Review of Economic Studies*, vol. 78, no. 1, pp. 176–198.

Pierre-Andre CHIAPPORI (1988), Rational household labor supply. *Econometrica*, vol. 56, no. 1, pp. 63–90.

Pierre-Andre CHIAPPORI (1992), Collective labor supply and welfare. *Journal of Political Economy*, vol. 100, no. 3, pp. 437–467.

Pierre-Andre CHIAPPORI & Ivar EKELAND (2006), The micro economics of group behavior: General characterization. *Journal of Economic Theory*, vol. 130, no. 1, pp. 1–26.

Pierre-Andre CHIAPPORI & Ivar EKELAND (2009), The Microeconomics of Efficient Group Behavior: Identification. *Econometrica*, vol. 77, no. 3, pp. 763–799.

Rahul DEB, Yuichi KITAMURA, John K.H. QUAH & Jörg STOYE (2017), Revealed Price Preference: Theory and Empirical Analysis. *Cowles Foundation Discussion Paper*, vol. 1, no. 2087, pp. 1–22.

Holger DETTE, Stefan HODERLEIN & Natalie NEUMEYER (2016), Testing multivariate economic restrictions using quantiles: the example of Slutsky negative semidefiniteness. *Journal of Econometrics*, vol. 191, no. 1, pp. 129–144.

Kirill EVDOKIMOV (2010), Identification and Estimation of a Nonparametric Panel Data Model with Unobserved Heterogeneity. *Princeton University: Discussion Paper*, pp. 1–62.

Vojtech FRANC, Vaclav HLAVAC & Mirko NAVARA (2005), Sequential coordinate-wise algorithm for the non-negative least squares problem. *Computer Analysis of Images and Patterns*, vol. 3691, pp. 407–414.

George-levi GAYLE & Andrew SHEPHARD (2016), Optimal Taxation, Marriage, Home Production, and Family Labour Supply.

Stefan HODERLEIN & Joerg STOYE (2014), Revealed Preferences in a Heterogeneous Population. *Review of Economics and Statistics*, vol. 96, no. 2, pp. 197–213.

Stefan HUBNER (2018), Reacher or Settler? Unobserved Distribution Factors and Preferences in the Collective Household Model. *Unpublished Manuscript*.

Bjoern JOHANSSON, Tommy ELFVING, Vladimir KOZLOV, Yair CENSOR, Per-Erik FORSSEN & Goesta GRANLUND (2006), The application of an oblique-projected Landweber method to a model of supervised learning. *Mathematical and Computer Modelling*, vol. 43, no. 7, pp. 892–909.

Yuichi KITAMURA & Joerg STOYE (2013), Nonparametric Analysis of Random Utility Models: Testing. *Cowles Foundation Discussion Paper*, vol. 1902, pp. 1–48.

Charles L. LAWSON & Richard J. HANSON (1995), Solving Least Squares Problems. In SIAM, classics i ed.

Daniel MCFADDEN (2005), Revealed stochastic preference: a synthesis. *Economic Theory*, vol. 26, pp. 245–264.

Daniel MCFADDEN & Marcel RICHTER (1991), Stochastic rationality and revealed stochastic preference. In Preferences, Uncertainty, and Optimality (edited by J. CHIPMAN, D. MCFADDEN & M.K. RICHTER.), Boulder: Westview Press, pp. 161–186.

Hal VARIAN (1982), The nonparametric approach to demand analysis. *Econometrica*, vol. 50, no. 1, pp. 945–73.

Frederic VERMEULEN (2006), Collective Models of Labor Supply with Nonconvex Budget Sets and Nonparticipation: A Calibration Approach. *Review of Economics of the Household*, vol. 4, no. 2, pp. 113–127.

A.1 Collective Axiom of Revealed Preference

Definition 1 (Collective Axiom of Revealed Preference, Cherchye, De Rock & Vermeulen (2007))**.** *Suppose that there exists a pair of utility functions* u^f *and*

u^m *that provide a collective rationalization of the set of observations* $\{(p_t; \tilde{x}_t^c, \tilde{x}_t^f, \tilde{x}_t^m) : \tilde{x}_t^c = \tilde{x}_t^f + \tilde{x}_t^m$
Then there exist preference relations[12] R_0^r *and* R^r *for each* $r \in \{c, m, f\}$ *such that:*

(i) *if* $\tilde{x}_s R_0^c \tilde{x}_t$, *then* $\tilde{x}_s R_0^f \tilde{x}_t$ *or* $\tilde{x}_s R_0^m \tilde{x}_t$

(ii) *if* $\tilde{x}_s R_0^r \tilde{x}_{s_1}, \tilde{x}_{s_1} R_0^r \tilde{x}_{s_2}, \ldots, \tilde{x}_{s_S} R_0^r \tilde{x}_t$ *then* $\tilde{x}_s R^r \tilde{x}_t$ *for* $r \in \{m, f\}$

(iii) *if* $\tilde{x}_s R_0^c \tilde{x}_t$ *and* $\tilde{x}_t R^r \tilde{x}_s$, *then* $\tilde{x}_s R_0^{r'} \tilde{x}_t$ *for* $r \neq r'$ *where* $r, r' \in \{m, f\}$

(iv) *if* $\tilde{x}_s R_0^c (\tilde{x}_{t_1} + \tilde{x}_{t_2})$ *and* $\tilde{x}_{t_1} R^r \tilde{x}_s$ *then* $\tilde{x}_s R_0^{r'} \tilde{x}_{t_2}$ *for* $r \neq r'$ *where* $r, r' \in \{m, f\}$.

(v) *if* $\tilde{x}_{s_1} R^f \tilde{x}_t$ *and* $\tilde{x}_{s_2} R^m \tilde{x}_t$ *then* $\neg (\tilde{x}_t R_0^c (\tilde{x}_{s_1} + \tilde{x}_{s_2}))$

(vi) *if* $\tilde{x}_s R^f \tilde{x}_t$ *and* $\tilde{x}_s R^m \tilde{x}_t$, *then* $\neg (\tilde{x}_t R_0^c \tilde{x}_s)$

where R^r *is defined as* $\tilde{x}_s R_0^r \tilde{x}_t$ *whenever* $p_s \tilde{x}_s^r \geqslant p_s \tilde{x}_t^r$ *and* R^r *is the transitive closure of* R_0^r *(Afriat, 1967; Varian, 1982).*

A.2 Proofs

Proof of Lemma 1. The equivalences between (i), (ii) and (iii)' are shown in McFadden & Richter (1991); McFadden (2005). Statement (iii)' referenced therein, differs from (iii) in that it additionally requires $\iota^T \nu = 1$. We now show that this is implied. It is easy to see that by construction of A for any solution of the quadratic problem we have $\eta = \pi$ and since $3 = \iota^T \pi = \iota^T A \nu = 3\iota^T \nu$ by construction of the 3-petite graph, we get $\iota^T \nu = 1$. Thus constraint $\nu \geqslant 0$ in is sufficient for η to be on the probability simplex.

It will be useful to write this problem with a *tightened* cone constraint indexed by $\underline{\nu}$. Let L be a lower diagonal matrix from the Cholesky decomposition $\Omega = LL^T$. Then we can rewrite the quadratic form (iii) as

$$\min_{\eta \in \{A\nu | \nu \geqslant \underline{\nu}\}} (\pi - \eta)^T LL^T (\pi - \eta).$$

Using $\eta = A\nu$ and introducing a slack variable $s \geqslant 0$ such that we can write $\nu = \underline{\nu} + s$ we obtain

$$\min_{\nu = \underline{\nu} + s, s \geqslant 0} (\pi - A(\underline{\nu} + s))^T LL^T (\pi - A(\underline{\nu} + s)).$$

This does not depend on ν but only on s and we can write it in the quadratic

[12] Note that R_0^c is just notation and not actually a preference relation, since household consumption is only a result of individual preferences.

form

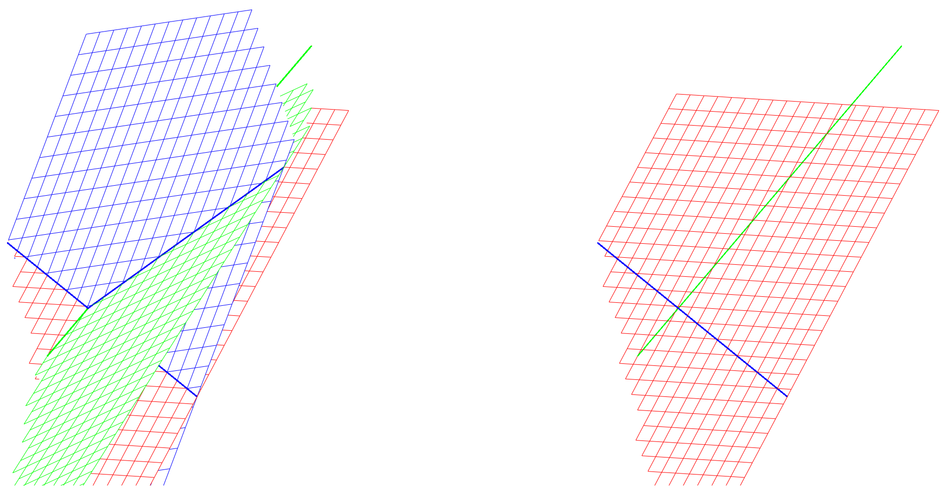
$$\min_{s \geqslant 0} \left\{ \frac{1}{2} s^T A^T \Omega A s - s^T A^T \Omega (\pi - A\underline{\nu}) \right\}.$$

Letting $H = A^T \Omega A$ and $f(\pi, \underline{\nu}) = -A^T \Omega(\pi - A\underline{\nu})$ we get a canonical form of a non-negative least squares problem, with gradient for iteration $\tau \geqslant 0$ defined as $\mu_\tau = H^T s_\tau + f(\pi, \underline{\nu})$. Johansson *et al.* (2006) show that component–wise projection $s_{\tau+1,j} = \max(0, s_{\tau,j} - \mu\tau, j d_j$ where $d = \text{diag}(H\iota)^{-1}$ and $j \in I^{|\mathfrak{X}^0|}$ referring to the j^{th} component of s will find the solution of the problem. □

A.3 A Minimal Example

To test the collective model using revealed preferences, at least three periods and three different goods are needed. This section discusses the strategy and dimensionality of our test in such a minimal setting.

Figure 2: Three intersecting budget sets $B_{\text{red}}, B_{\text{blue}}, B_{\text{green}}$ with three goods



Note: Example of a three-good economy with three price-regimes characterizing budgets B_t where $t \in \left\{\underline{b}\text{lue}, \underline{r}\text{ed}, \underline{g}\text{reen}\right\} = I_T$. In the figure on the right hand side the green and blue budgets are removed and only the lines in which they intersect with the remaining red budget are plotted.

We can make revealed preference statements whenever some bundle of goods was chosen, but a different less expensive one would have been affordable. In Figure 2, the colors of the budget sets are defined such that whenever a household chose one of the four regions on a budget set that was above another budget set, then the good corresponding to this color is revealed preferred to the good corresponding to the color of the other budget set. For example if a single female household picks one of the lower quadrants of the red budget such that $p_{\text{blue}} x_{\text{red}} \leqslant p_{\text{red}} x_{\text{red}}$, we can say that $x^{\text{red}}\ R^f\ x^{\text{blue}}$.

To clarify the construction of preference types from continuous data, we will now consider an example of a hypothetical household match, faced with the respective budgets. For this, let us assume we observe a single female consuming x_r^f, x_b^f and x_g^f, a single male consuming x_r^m, x_b^m and x_g^m and consider them to be matched into a couple of type equivalent to one consuming x_r^c, x_b^c and x_g^c. We normalize household endowment to one. Their consumption satisfies the following inequalities which contain all the necessary information to characterise the matched household in terms of preference relations which can be checked against the Collective Axiom of Revealed Preference.

$$\begin{array}{lll|ll|ll} p_b x_r^c \geqslant 1, & p_g x_r^c \geqslant 1, & p_r(x_b^c + x_g^c) \geqslant 1 & p_b x_r^m \leqslant 1, & p_g x_r^m \geqslant 1, & p_b x_r^f \geqslant 1, & p_g x_r^f \leqslant 1 \\ p_r x_b^c \geqslant 1, & p_g x_b^c \geqslant 1, & p_b(x_r^c + x_g^c) \geqslant 1 & p_r x_b^m \geqslant 1, & p_g x_b^m \leqslant 1, & p_r x_b^f \geqslant 1, & p_g x_b^f \geqslant 1 \\ p_r x_g^c \leqslant 1, & p_b x_g^c \geqslant 1, & p_g(x_r^c + x_b^c) \leqslant 1 & p_r x_g^m \geqslant 1, & p_b x_g^m \geqslant 1, & p_r x_g^f \geqslant 1, & p_b x_g^f \geqslant 1 \end{array}.$$

If f and m were to be matched together, the inequalities restricting x^c represent one potential joint consumption type, in which case the graphs of the respective preference relations are

$$R_0 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad R_0^m = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \qquad R_0^f = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

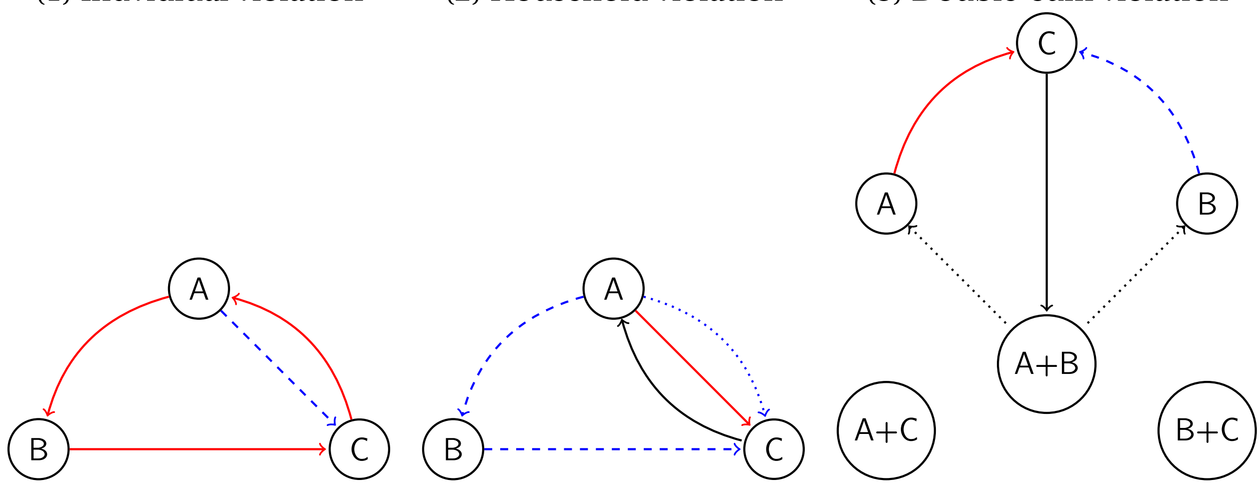
with transitive closures for both individuals:

$$R^m = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \qquad R^f = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Table 3 shows different violations of the collective axiom. Situation (1) is a trivial violation of individual SARP by f. In the example above we have situation (2), in which the preference relation R_0^m is one of a person who prefers good A over good B and good B over good C. Thus he must also prefer good

Table 3: Violations of the Collective Axiom

(1) Indvidiual violation (2) Household violation (3) Double-sum violation



Note: Nodes refer to different consumption bundles. the red solid line to the wife's preference relation R_0^f, the r ed dotted line to the implied transitive closure R^f, the blue dashed line to the husband's preference relation R_0^m, and the black solid line to the household "preference" relation.

A over good C by transitivity, for which there is no contradicting revelation of preferences. Hence, this person is rational. The preference relation of R_0^f also represents a rational person, who prefers good A over good C. Note that this implies that both individuals prefer good A over good C, but aggregate household consumption represented by R_0^c revealed that the household chose good C over good A; a violation of the Collective Axiom of Revealed Preference. This households is also irrational due to a violation of type (3), in which the household could have consumed both A and B, but chose to consume only C instead, making both individuals worse off.

In our setting, the cardinality of $\mathcal{X}^r$ is $|\mathcal{X}^r| = 4^3 = 64$ for $r \in \{m, f\}$ representing single males and single females, respectively. For couples, we have to evaluate inequalities for double-sums according to Definition 1 (iv) & (v), for which we have 2^κ different possibilities with $\kappa = \frac{1}{2}T(T-1) = 3$ which results in $4^3 * 2^3 = 512$ choices. In total, we have thus $|\mathfrak{X}| = 512 * 64 * 64 = 2{,}097{,}152$ matched household types. Applying a revealed preference test for this universe of types and removing the ones that violate the collective axiom under the preference stability assumption, we end up with $|\mathfrak{X}^0| = 2{,}996$ matched household types.

$|\mathfrak{X}^{\text{collective}}| = 475{,}136$ are consistent with the collective axiom based on the necessary conditions using only aggregate household consumption data. This leaves us with about 22.7% collectively rational types. From this, we should not necessarily conclude a restrictive nature of the collective model since for a given range of budget planes only a subset of the total choice set would actually be feasible (e.g. have positive demands). Hence we obtain a matrix A with $512 + 64 + 64 = 640$ rows and 2,996 columns representing rational types under the stable preference assumption. The vector π is a vector of choice probabilities of the population of the same dimension: 640. While this might seem high-dimensional we note that this matrix is very sparse. In fact, A only has $3|\mathfrak{X}^0| = 3 * 2{,}996$ non-zero items.

A.4 Simulations

In this section, we investigate the properties of our proposed test in a simulation setting. In particular, we are interested in how much power it has to detect a violation of the stable preference assumption and whether or not it has a correct proportion of false positives. Since specifying a parametric continuous demand system requires at least five goods to impose the SNR(S-1) condition on the Slutsky matrix and distinguish the collective model from the unitary model, we will not sample continuous demands as functions of prices and individual budget constraints, but rather draw our sample directly from the discrete choice space.[13] This should be interpreted as a continuous uniform distribution of choices on different budget planes, where the relative prices are such that the partitions of the budget planes are of equal size. Recall that we test this against the set of households which are consistent with the necessary conditions of the collective axioms based on aggregate consumption but not consistent when single data and the stable preference assumption is added. This set is denoted by $\mathfrak{X}^1$ and we have $\mathfrak{X}^{\text{collective}} = \mathfrak{X}^0 \cup \mathfrak{X}^1$. If we reject the null hypothesis that both the collective axiom and the stable preference

[13] A revealed preference based setting allows us to test the restrictions of the model with only three goods (Cherchye, De Rock & Vermeulen, 2007), whereas Browning & Chiappori (1998) need five goods.

assumption holds, by excluding all irrational matches $\mathfrak{X} \setminus \mathfrak{X}^{\text{collective}}$, we must conclude that the stable preference assumption does not hold. To control the proportion of households for whom this is the case (our data generating process) we introduce the parameter p which specifies the probability that a particular choice is both collectively rational and satisfies the stable preference assumption $p := \mathbf{P}(x \in \mathfrak{X}^0)$.[14] By only considering collectively rational choices in our simulations we thus have $1-p = \mathbf{P}(x \notin \mathfrak{X}^0) = \mathbf{P}(x \in \mathfrak{X}^1)$ by construction.

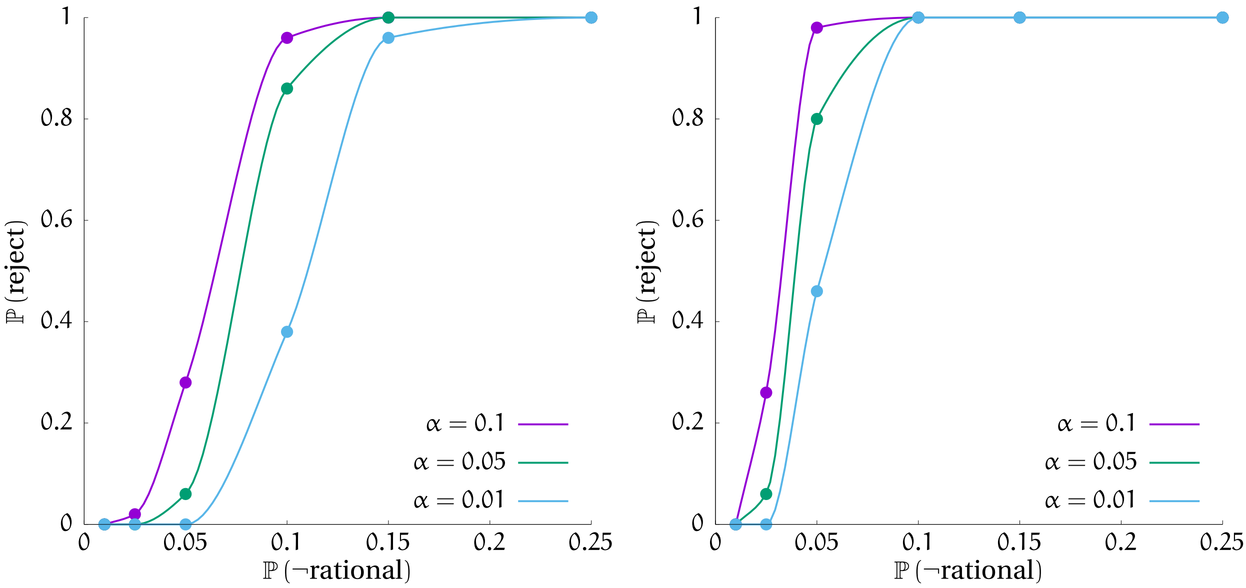
Our simulation setting is as follows. We consider $S = 100$ samples of size $\underline{N} \in \{500, 1000, 2000\}$ where $\underline{N} = N_f = N_m = N_c$ such that $N = 3\underline{N}$ in a minimal setting with $T = 3$ periods which we construct by drawing $\lfloor \underline{N}p \rfloor$ indices from the space of collectively rational matches $\mathfrak{X}^0$ for which the stable preference assumption holds and $\lceil \underline{N}(1-p) \rceil$ indices from the space of collectively rational types $\mathfrak{X}^1$ which does not satisfy the assumption. Based on a sample of matches, we then calculate the choice probabilities $\widehat{\pi}$ accordingly. For estimation, we only use the marginal distribution of choices of each sample of household compositions and draw $B = 100$ samples from the respective empirical distributions (i.e. with replacement) to calculate $\pi^b_{\tau_N}$ and estimate the empirical distribution of the test statistic $\mathcal{J}^{\tau_N}_{N,b}$. These simulations are repeated for $p \in \{0.75, 0.85, 0.9, 0.95, 0.975, 0.99, 1.00\}$.

Figure 3 shows the power of our test against the non-stable preference alternative as a function of p, with sample-size $\underline{N} = 500$ for the left-hand side graph, and $\underline{N} = 1000$ for the right-hand side graph, respectively. We use monotone cubic splines to interpolate between the actual simulation results, which are marked as solid dots. To be more precise, the respective functions refer to sample rejection frequencies using the rejection rule $J \mapsto \mathbb{1}\left\{J > \widehat{F^{-1}_{\mathcal{J}_N}}(1-\alpha)\right\}$ for $\alpha \in \{0.01, 0.05, 0.10\}$. In addition to this, we also observe that as $\underline{N}$ increases the power of our test improves and is able to correctly reject the hypothesis of a collectively rational population already at small proportions p.

The intercepts of these functions should be interpreted as the proportion of false positives (type I errors) since they correspond to the case where everyone is rational. One might expect that for a correctly sized test the empirical

[14]This rationality parameter is similar as for example λ in Dette, Hoderlein & Neumeyer (2016) which specifies the population's deviation from Slutsky symmetry.

Figure 3: Power function for $N = 1{,}500$ (l.h.s) and $N = 3{,}000$ (r.h.s.)

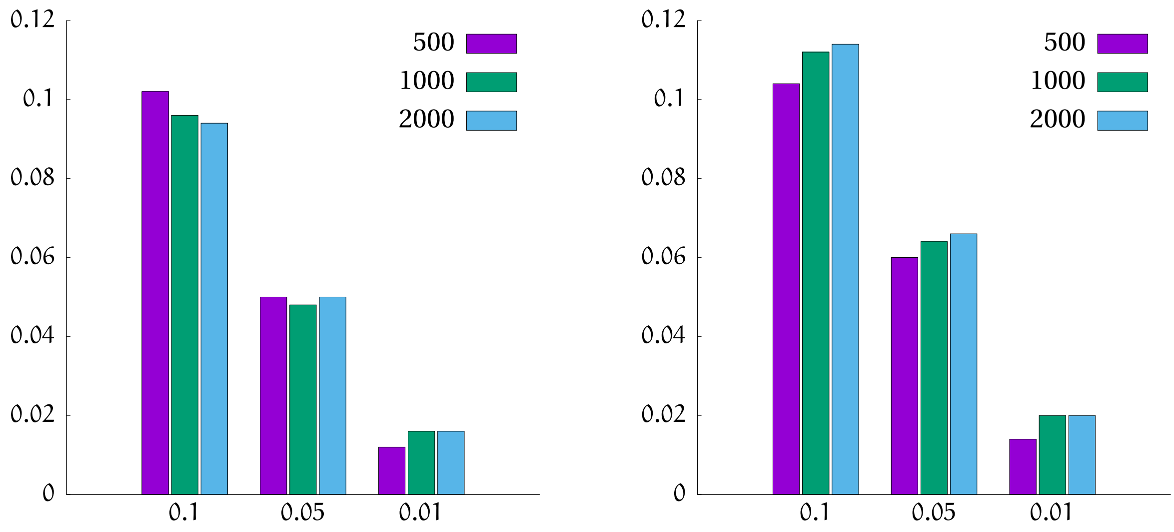


rejection frequencies should tend to α. However, given our partial identification procedure we have a composite null hypothesis, *i.e.* the probability of a type I error should be at most α as defined in equation (3). To see this note that every vector of "true" choice frequencies denoted by π_0 lying in the interior of the cone will have projection residuals of length zero. Bootstrapping out of $\widehat{\pi}$ which tends to π_0 using the usual regularity properties could then lead to a confidence interval which is always entirely in the interior of the cone and we would never wrongly reject the null hypothesis. This also implies that in such a case our bootstrap distribution is degenerate and has mass one at point zero.

In our Monte Carlo setting and the case where $p = 1.0$, we randomly select types from the type-space $\mathfrak{X}^0$, satisfying collective rationality. Thus the "true" parameter vector ν_0 is assumed to have a uniform distribution over the probability simplex and the worst-case, namely to get a ν such that $\pi_0 = A\nu$ is on the boundary of the cone with respect to any of its dimensions, occurs with measure zero.

Thus, in order to evaluate whether the size of our test is correct under the test's minimax strategy, we have to construct a worst case. For this, note that the test is constructed in a way that considers hypothetical types by taking combinations of possible household choice behaviour per price regime over

Figure 4: Type I error for $n_0 = 5$ (l.h.s) and $n_0 = 2$ (r.h.s.) worst-case matches



a range of price regimes. To fix notation, we will call two collectively rational matches *similar* if there is at least one element in the product space spanned by these two matches which is an element of the space of collectively rational matches that do not satisfy the stable preference hypothesis. We will then construct worst cases by specifying a distribution over n_0 such similar matches. To make sure that our π_0 is on the boundary of the cone in all dimensions, *i.e.* on the cusp, we shift the cone by manually controlling the tightening parameter τ_N according to this distribution. Figure 4 shows simulation results for two such worst case scenarios with 5 similar matches and 2 similar matches, respectively.

While both are asymptotically valid from a theoretical point of view, it is not surprising that for finite samples the size for the case with a larger number of worst case matches behaves worse than the case in which there are fewer worst case matches. Since the properties of the test are based on an asymptotic argument, we should see the empirical frequency of false positives tending to the respective α which define the rejection rules and are plotted on the x-axis. The results are what one would expect, with all sample sizes being reasonably accurate. Since in a well-behaved test, false-positives are by definition rather rare events, in order to minimize simulation uncertainty, we increased the number of Monte Carlo repetitions to $S = 500$ and the number of bootstrap repetitions to $B = 200$, which greatly increased computational complexity due to the high dimensionality of the testing problem.

A.5 Descriptive Statistics

Table 4: RLMS: Monthly Consumption

		Dairy			Bread			Meat		
Year	N	Mean	IQR	P	Mean	IQR	P	Mean	IQR	P
2000	1506	81.7	120.4	121.1	113.2	102.8	116.5	322.0	439.7	128.3
2005	1601	222.3	277.8	110.5	199.3	171.8	103.0	1003.5	1175.8	118.6
2010	2839	475.7	492.7	116.7	303.3	270.8	107.6	1862.1	1926.0	105.3
2011	2983	520.3	544.8	106.3	317.3	270.9	108.9	2165.8	2154.7	109.2
2012	3154	551.0	550.5	104.4	330.2	291.7	112.0	2284.8	2315.0	108.3
2013	3076	617.7	622.9	113.1	352.9	287.1	108.0	2366.2	2399.8	97.0
2014	2516	695.3	675.8	114.4	372.9	323.0	107.5	2805.6	2847.5	102.1

Note: Descriptive statistics of the Russian Longitudinal Monitoring Survey (RLMS) reporting mean, interquantile range (IQR) and price index P for monthly consumption of composite goods in a given year. All quantities are inflated to 2014 prices and denoted in local currency (Russian Ruble). Goods are aggregated to composite good categories as follows. Dairy: Canned/powdered milk, fresh milk, sour milk products and sour cream; Bread: White (wheat) bread and black (rye) bread; Meat: Canned meat, beef/veal, lamb/goat, pork, giblets, poultry, lard, sausage and semi-prepared meat products

Table 5: ECPF: Weekly Consumption

Year	N	Clothing			Food out			Nondurables		
		Mean	IQR	P	Mean	IQR	P	Mean	IQR	P
1985	65	1284.4	1406.4	165.7	940.9	899.7	174.3	35.1	43.8	150.6
1986	95	1334.8	1545.3	191.6	927.3	1128.1	224.7	28.8	47.6	164.5
1987	288	1743.3	1897.1	174.2	1054.7	1244.8	191.5	37.0	53.8	157.2
1988	195	1537.6	1831.0	158.1	1036.2	1400.9	160.0	42.6	53.3	145.8
1989	225	2253.9	2344.4	134.3	1446.3	1685.6	139.6	57.2	70.5	140.7
1990	205	2289.9	2565.5	106.6	1636.4	2152.1	112.2	41.0	53.2	101.9
1991	210	2255.2	2398.5	183.5	1852.4	2229.4	208.7	52.5	66.7	160.8
1992	202	2652.5	2795.1	154.5	1852.6	1957.9	154.5	69.1	85.4	144.0
1993	185	2823.0	2471.3	112.4	2386.2	3022.2	121.0	75.7	80.8	114.5
1994	210	2102.8	2471.0	106.4	2322.7	2730.5	111.2	79.8	97.4	102.9
1995	194	2186.9	2287.9	113.6	2068.9	2187.8	122.2	117.8	118.3	114.4
1996	199	2397.4	2595.2	126.5	2761.4	3230.4	126.6	107.5	129.6	129.5

Note: Descriptive statistics of the Spanish Continuous Family Expenditure Survey (Encuesta Continua de Presupuestos Familiares) reporting mean, interquantile range (IQR) and price index P for weekly consumption of goods in a given quarter. We only report descirptive statistics of the first quarter of a given year. All quantities are normalized to arbitrary units using the price indices P.