A positive theory of cooperative games: The logit core and its variants

Friedel Bolle and Yves Breitmoser and Philipp E. Otto

EUV Frankfurt (Oder)

19. August 2011

A positive theory of cooperative games: The logit core and its variants*

Friedel Bolle[†] Yves Breitmoser Philipp E. Otto

EUV Frankfurt (Oder) EUV Frankfurt (Oder) EUV Frankfurt (Oder)

August 19, 2011

Abstract

This paper proposes two generalization of the core and evaluates them on experimental data of assignment games (workers and firms negotiate wages and matching). The generalizations proposed allow for social utility components (e.g. altruism) and random utility components (e.g. logistic perturbations). These generalizations are well-established in analyses of non-cooperative games, and they prove to be both descriptive and predictive in the assignment games analyzed here. The "logit core" allows us to define a "stochastically more stable" relation on the outcome set, which has intuitive implications, and it fits better than alternative approaches such as random behavior cores and regression modeling.

JEL classification: C71, C90, D64

Keywords: cooperative games, core, random utility, social preferences, laboratory experiment

*The financial support of the DFG (project no. $BO-747/11-1$) is gratefully acknowledged.

[†]Corresponding author. Postfach 1786, 15207 Frankfurt (Oder), Germany, Email: bolle@euv-frankfurt-o.de, Phone/Fax: +3355534 2289/2390

1 Introduction

In this paper, we consider the assignment problem originally defined by Koopmans and Beckmann (1957) and Shapley and Shubik (1972). There are sellers and buyers of indivisible goods (e.g. firms and workers in labor markets) who generate a surplus when they match. A worker (seller) can match with only one firm (buyer) and vice versa. Firms and workers can share the surplus they generate however they will, i.e. utility is transferable within matches. The latter assumption contrasts with other models of matching markets, e.g. between students and universities or men and women, where utility is non-transferable (following Gale and Shapley, 1962). The resulting "assignment game" is the cooperative game based on the assignment problem where firms and workers negotiate overall matching and individual surplus allocation.

Theoretically, the assignment game is well-understood. The canonical solution concept is the core and it has many "nice" properties. It is non-empty and characterized as the solution set of a linear programming problem (Koopmans and Beckmann, 1957), it is a polytope with the form of a 45°-lattice (Quint, 1991a), it satisfies the so-called CoMa property (Hamers et al., 2002), it has been axiomatized (Toda, 2005), and there exist mechanisms implementing core points (Pérez-Castrillo and Sotomayor, 2002; Halaburda, 2010). Some of these properties do not generalize to m-sided markets (Quint, 1991b), but they tend to generalize to one-sided matching (Quint, 1996) and multiple-partners games (Sotomayor, 1999), and overall, the assignment game is considered a standard model of labor markets (Crawford and Knoer, 1981; Kelso Jr and Crawford, 1982) and other markets with indivisible goods (Roth, 1985).

Empirically, however, there are various open questions. Most existing experimental research focuses on the efficiencies of mechanisms for matching markets (see Olson and Porter, 1994; Nalbantian and Schotter, 1995, for mechanisms with transfers, and Kagel and Roth, 2000; Chen and Sönmez, 2002, 2006; Pais and Pintér, 2008, for mechanisms without transfers), which is important to evaluate mechanisms and the underlying assumptions. By their very design, however, these experiments cannot inform us on the empirical relevance of the core in the standard worker-firm problem. The only experimental analyses of assignment games without mechanisms, where negotiations and rematching are unrestricted, seem to be Tenbrunsel et al. (1999) and Otto and

Bolle (2011). Their main findings are that the core fits comparably poorly, as many of the experimental observations are not in the core, and that the fit can be improved by weakening the assumption of rationality on the side of the players. In a first step of weakening rationality, the "equal division core" proposed by Selten (1972) fits better than the core,[1] and secondly, the even simpler "equal split solution" (where players are satisfied if their payoff does not differ too much from the equal split of the productivity in their current match) improves upon the equal division core. Thus, subjects do not seem to consider the possibility of alternative matches at all.

These findings relate loosely to the level-k theory of play in non-cooperative games (Stahl and Wilson, 1995; Camerer et al., 2004; Costa-Gomes et al., 2009). There, level-1 players believe the opponents are non-strategic, level-2 believe the opponents are level 1, and so on. In assignment games, the equal split solution assumes that alternative matches are impossible, and the equal division core assumes that the equal split results in alternative matches. These concepts relate to level 1 and level 2 in non-cooperative games, and their empirical fit raises the question if another concept that proved highly relevant in non-cooperative games, *social preferences*, would affect reasoning in cooperative games, too. This seems intuitive, since emotions such as altruism or inequity aversion are widespread, and the evidence for their existence in non-cooperative games is overwhelming (for a survey, see e.g. Camerer, 2003), but for cooperative games, this explanation has not yet been explored. We will explore it and provide evidence that subjects seem to have interdependent preferences, namely "spiteful" preferences as a result of which they bargain particularly aggressively.

The main purpose of our paper concerns the analysis of a more fundamental concern, however. Why do subjects deviate from the core and its variants in the first place? Understanding the sources of deviations is important for various reasons. First, an explanation of deviations is required to predict the overall distribution of outcomes, i.e. to apply the model empirically. Second, a structurally complete econometric model of play in cooperative games enables us to apply techniques such as maximum likelihood

[1] Formal definitions follow below. Briefly, an outcome is in the core if no worker-firm pair can rematch profitably by allocating their surplus in any way. The equal-division core weakens that matches be stable with respect to any alternative surplus allocation, by requiring stability only with respect to the equal division of surpluses in alternative matches. Thus, subjects do not have to be able to evaluate all alternative allocations.

estimation to obtain efficient estimates of model parameters (such as level of reasoning or altruism coefficients). Third, it allows us to define a *gradual* measure for the stability of outcomes, and thus to generalize the binary measure currently used. This yields a "stochastically more stable" relation between outcomes, which allows us to investigate if allocations in the interior of the core are more stable than those close to its boundary, and to define entirely new concepts merging ideas from those discussed above (as we do below). Finally, the sources of deviations from rationality in choice under risk and in non-cooperative games have been analyzed in much detail (see e.g. Wilcox, 2008, 2011, Conte et al., 2011, for choice under risk, and McKelvey and Palfrey, 1995, Weizsäcker, 2003, Goeree and Holt, 2004, for non-cooperative games), and thus a corresponding analysis of cooperative games clarifies if there are fundamental differences between non-cooperative games and cooperative games at level of say utility functions and rationality.

Our results can be summarized as follows. We find that the source of deviations from the core is best modeled as a logistic random utility component (the resulting logit core relates naturally to the logit equilibrium defined by McKelvey and Palfrey, 1995). The random utility models fit highly significantly better than a random behavior model, where the outcome is in the core with probability $1-\varepsilon$ and outside of it with probability ε, and it also fits better than regression modeling. We determine both the descriptive accuracy and the predictive accuracy, similar to Hey et al. (2010) for choice under risk, and show that the random utility models fit robustly. Further, we find that preferences have spiteful components also after accounting for random utility, i.e. subjects bargaining very competitively. The best fitting concept (both descriptively and predictively) unifies the equal division core (level 2) and the equal split solution (level 1). These result can be applied readily in further experimental and empirical analyses.

Section 2 defines assignment games and the core, and it describes the experiment. Section 3 reviews the basic experimental results (originally described in Otto and Bolle, 2011). Section 4 introduces social preferences into the cooperative game and Section 5 examines their fit. Section 6 introduces random utility into the cooperative game, and Section 7 discusses the model fit and estimates. Section 8 concludes.

2 Basic definitions and experimental design

Assignment games and the core

Let W be a finite, non-empty set of "workers" and F be a finite, non-empty set of "firms." The productivity of the potential matches $(w, f) \in W \times F$ between workers and firms is denoted as $C \in \mathbb{R}_+^{W \times F}$, i.e. $C_{w,f}$ is the value if w and f match. The allocation of their value $C_{w,f}$ is to be negotiated between w and f. Players that are unmatched obtain zero payoff. The outcome of an assignment game is a payoff profile $(x_i)_{i \in W \cup F}$.

The *core* contains all outcomes where no subset of players can increase their payoffs by rematching. An outcome $(x_i) \in \mathbb{R}_+^N$, $N = W \cup F$, is in the core if (and only if) it is feasible and $x_w + x_f \geq C_{w,f}$ for all $w \in W, f \in F$. All core outcomes are socially efficient, i.e. they maximize the productivity aggregated over all matches. Koopmans and Beckmann (1957) and Shapley and Shubik (1972) show that the core is generally non-empty (in assignment games) and that transfers between matches are neither made nor required to sustain core allocations.[2] Solymosi and Raghavan (2001) provide necessary and sufficient conditions for the core to be stable in the sense of von Neumann-Morgenstern, and these conditions will be satisfied in our experimental games. Driessen (1998) shows that the kernel is included in the core of assignment games, and Núñez and Rafels (2003) derive a similar result for the τ-value.

The treatments

All of the games in our experiment are 2×2 assignment games, and the productivity matrices for all the treatments $T1 \ldots T6$ are provided in Table 1. In all treatments, $C_{1,1} \leq C_{1,2} \leq C_{2,1} < C_{2,2}$ applies. There are assignment games that cannot be rearranged (by relabeling workers and firms) to satisfy this constraint.[3] In the present paper we focus on this class of assignment games, as it allows us to distinguish between players that are relatively strong (W_2 and F_2) and relatively weak (W_1 and F_1).

[2] This is not the case for most alternative solution concepts in assignment games. For example, nucleolus, Shapley Value, and Stable Sets (the von Neumann-Morgenstern Solution) of the Assignment Game require the possibility of transfers between matches.

[3] Alternatively, either $C_{1,2} < C_{1,1} < C_{2,1} < C_{2,2}$ or $C_{1,1} < C_{1,2} < C_{2,2} < C_{2,1}$ may apply.

Table 1: Productivities of matches in the six experimental treatments

		$T1$		$T2$		$T3$		$T4$		$T5$		$T6$	
$C_{1,1}$	$C_{1,2}$	$\underline{280}$	400	$\underline{280}$	280	280	460	160	$\underline{400}$	160	$\underline{460}$	280	400
$C_{2,1}$	$C_{2,2}$	400	$\underline{640}$	520	$\underline{640}$	460	640	$\underline{520}$	640	$\underline{460}$	640	520	640

Note: $C_{w,f}$ is the productivity of the match $(w, f) \in W \times F$. The pairings in the socially efficient matching are underlined if unique. In $T3$ and $T6$, both matchings are efficient.

For, $C_{1,1} < C_{2,1}$ and $C_{1,2} < C_{2,2}$ implies that W_2 always has a better bargaining position than W_1, and similarly $C_{1,1} \leq C_{1,2}$ and $C_{2,1} < C_{2,2}$ implies that F_2 is stronger than F_1.

In 2×2 games, the players can match in either of two ways. The matching $\{(W_1, F_1), (W_2, F_2)\}$ will be called "A-matching," and $\{(W_1, F_2), (W_2, F_1)\}$ will be called "B-matching." In $T1$ and $T2$, A-matching is efficient, in $T4$ and $T5$, B-matching is efficient, and in $T3$ and $T6$, both matchings are efficient. In the latter case, the core is degenerate, i.e. it has zero volume in the outcome space. Otherwise, its volume is positive. For each of these efficiency conditions, we distinguish whether the productivity matrix is symmetric ($C_{1,2} = C_{2,1}$) or asymmetric ($C_{1,2} < C_{2,1}$). Thus, we obtain a 3×2 experimental design in aggregate and cover all relevant scenarios.

The symmetry condition $C_{1,2} = C_{2,1}$ is important, as it induces symmetry between W_1 and F_1 on the one hand and W_2 and F_2 on the other. Alternatively, if these firms match, i.e. if $\{(W_1, F_1), (W_2, F_2)\}$, under asymmetry $C_{1,2} < C_{2,1}$, then the possible blocking coalition (W_2, F_1) has higher productivity than (W_1, F_2), and thus W_1 is weaker than F_1 and F_2 is weaker than W_2. In this case, unequal splits of the productivities $C_{1,1}$ and $C_{2,2}$ are reasonable. In the corresponding treatment $T2$, where both asymmetry and efficiency of A-matching obtains, the equal split of incomes is not in the core. Thus, the experiment is designed to distinguish equal splits of incomes and core allocations.

Logistics

The experiment consists of 28 sessions, and in each session, eight subjects played the six assignment games defined above in random order. For each game, the subjects were randomly divided into two groups of 2×2 subjects each (= two markets). The subjects then negotiated matching and surplus allocation, for up to ten minutes. During these ten minutes, preliminary contracts were concluded and superseded by new preliminary contracts. Two players with a standing contract between them were not able to renegotiate their contract. New contracts were negotiable only with the other potential partner. At the end of the game, the standing contracts became binding. The subjects were paid according to the sum of their earnings in all of their final contracts.

The order of the treatments and the individual allocation to positions was randomized over the sessions. Every subject was allocated to a worker position three times and to a firm position three times. No subject interacted with the same co-participant in more than three of the six games. Every subject assumed each position W_1, W_2, F_1, and F_2 at least once. No subject participated in more than one experiment. In total, 224 subjects took part in the study. A session lasted about 1.5 hours, in all settings, and the average payment was € 13.41 in *Class*, € 11.75 in *Lab*, and € 11.48 in *Lab-info*. Payments were calculated purely from the negotiated contracts and no show-up fee was paid. The experimental instructions are available as supplementary material.

Finally, to examine whether the negotiation outcomes (and the underlying fairness benchmarks) depend on the circumstances of the negotiations, we varied these circumstances, i.e. setting (classroom versus laboratory) and the information of the subjects (private versus complete). If the outcomes depend significantly on these "circumstances," then a cooperative solution concept such as the core (which ignores the circumstances by definition) would be infeasible. The circumstances considered in the experiment are as follows.

- *"Class":* 10 sessions where subjects negotiate *face to face* with explicit information only about the productivities of their own matches.
- *"Lab":* 10 sessions where subjects negotiate *anonymously* via computer terminals with explicit information only about the productivities of their own matches.
- *"Lab-info":* 8 sessions where subjects negotiate *anonymously* via computer terminals with explicit information about the productivities of all possible matches.

In the laboratory, the subjects negotiated via computer terminals by making and accepting offers from the other players in specific boxes on the screen. In the classroom setting, the firms were assigned specific tables and workers approached them to

Table 2: Average wages (w_1, w_2) in the various treatments

		$T1$	$T2$	$T3$	$T4$	$T5$	$T6$
Li	*A*	(142,307)	(153,348)	(140,350)	(86,337)	(67,362)	(133,375)
	B	(313,360)	(118,237)	(220,258)	(120,297)	(143,266)	(203,347)
L	*A*	(131,340)	(133,327)	(150,340)	(72,326)	(85,331)	(133,334)
	B	(188,222)	(180,270)	(207,263)	(201,347)	(145,245)	(176,293)
C	*A*	(160,328)	(143,322)	(145,318)	(80,332)	(83,303)	(149,329)
	B	(188,239)	(158,319)	(199,249)	(167,331)	(200,299)	(160,324)

Legend: *Li* $\hat{=}$ Lab-info, *L* $\hat{=}$ Lab, *C* $\hat{=}$ Class; *A* $\hat{=}$ *A*-matching, *B* $\hat{=}$ *B*-matching

negotiate terms. In *Lab-info* the whole payoff matrix was provided for each treatment. In the two other cases (*Class* and *Lab*) only the productivities of the matches involving oneself were shown.

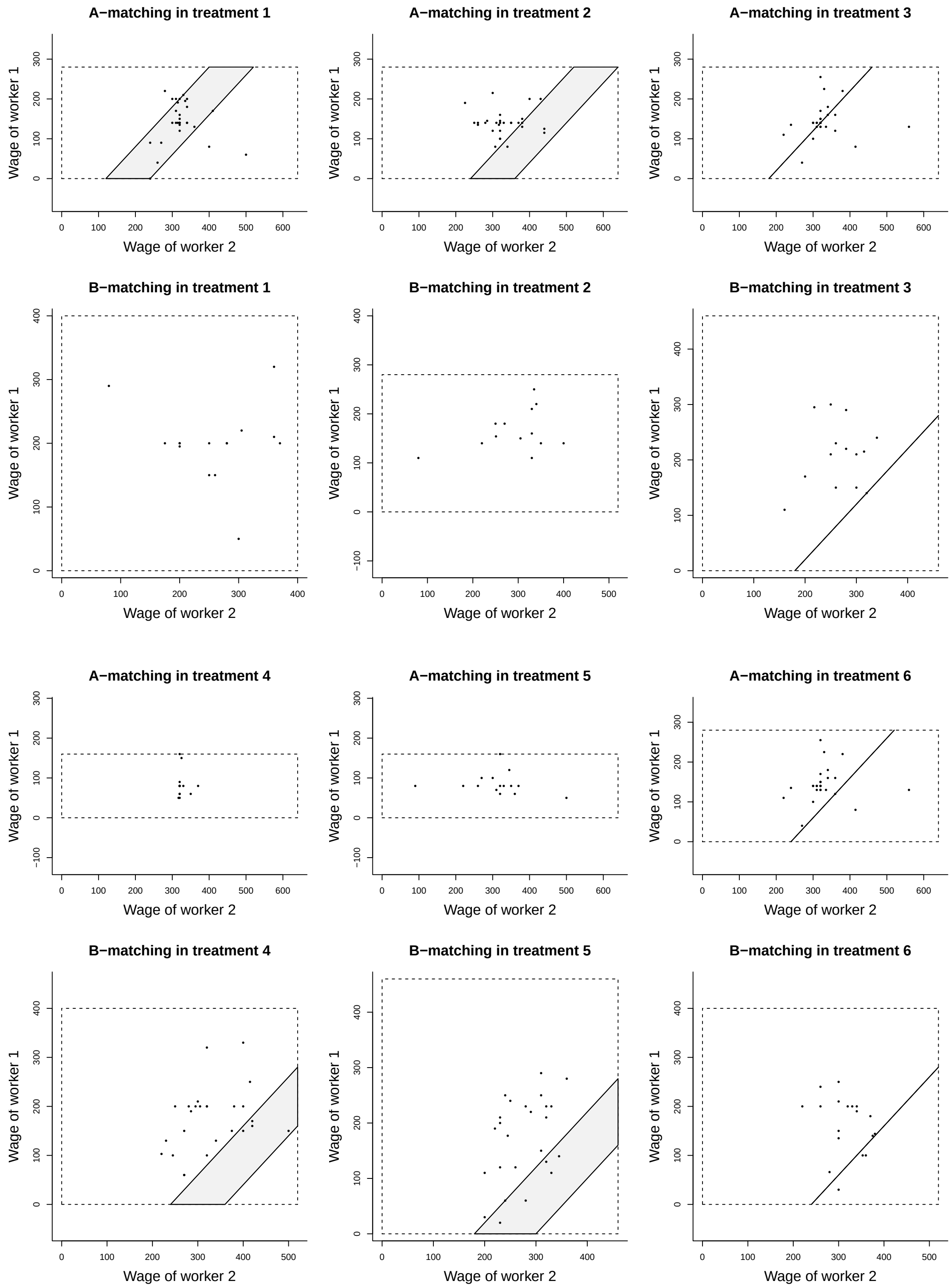
3 Basic results

The average wages negotiated in the various conditions are reported in Table 2 and Figure 1 plots all outcomes in relation to the core. A descriptive analysis of the experimental data is provided in Otto and Bolle (2011) and therefore this part is kept brief here. Otto and Bolle discuss the data in relation to basic bargaining theories such as the core, ε-core, Nash bargaining, Equal Split, ε-Equal Split, and Selten's (1972) Equal Share analysis. Their main finding is that the ε-Equal Split fits better than the other theories in explaining the results.

If we require stability as underlying the core or related concepts and assume that the negotiation process efficiently reveals productivities of all matches, then the three basic treatment designs *Class*, *Lab*, and *Lab-info* should be payoff equivalent. This hypothesis is confirmed in the sense that equality is not rejected significantly.[4] That is, the aspects of the game that the core neglects by definition (such as information and implementation) are empirically insignificant, and hence a cooperative approach

[4]In pairwise t-tests, the only significant difference between the means of wages (at $\alpha = .05$) can be found between the wages of W_2 in *Lab* and *Class* in treatment $T5$ under *B*-matching ($245 \neq 299$ at $\alpha = .05$). Only the variances of the wages differ slightly. They are highest in *Lab-info* (with overall standard deviation of w_1 being $\hat{\sigma}_{w_1} = 72.9$), intermediate in *Lab* ($\hat{\sigma}_{w_1} = 63.6$) and lowest in *Class* ($\hat{\sigma}_{w_1} = 48.1$).

Figure 1: The experimental data in relation to the core. (Note that if *B*-matching is inefficient, then the core predicts *A*-matching, and vice versa.)



is applicable indeed.

Overall, incomplete matching has been observed in 12% of the games. Incomplete matching is not stable unless we allow for highly spiteful preferences. The occasional incompletions resulted from last-second rematching under the 10-minute time line for the negotiations. This implies that the incomplete matches are likely not stable on their own. We will therefore focus on explaining the results of the 88% of the negotiations that resulted in complete (and stable) matching.

4 The generalized core and its variants

The empirical deviations of the observations from the core seem to be systematic, as the observations are not symmetric around the core (see Figure 1). Two possible explanations for such systematic deviations are that subjects have social preferences and that the stability requirements of the core are too strong or too computationally complex. We begin with formalizing social preferences.

By our experimental design, the players do not know the payoff allocation between the players in the other match. They may try to infer their payoffs during negotiations, but this inference will necessarily be incomplete. The following generalization therefore assumes that i's utility depends on the values that he can explicitly observe—the own payoff, the partner's payoff, and the partner's identity—but further generalization requires only minor adaption of the notation. To define the utility formally, define the set of players $N := W \cup F$, and additionally $N_i = F \cup \{\emptyset\}$ if $i \in W$ as well as $N_i = W \cup \{\emptyset\}$ if $i \in F$. Thus, N_i contains the potential partners of i, where "$\emptyset$" indicates that i remains single. The utility of $i \in N$ is $U_i(x_i, x_j, j) : \mathbb{R}^2_+ \times N_i \to \mathbb{R}$. We limit altruism and spite in U_i as follows.

Assumption 4.1. For all $i \in N$ and $j \in N_i$,

(i) $\partial U_i/\partial x_i - \partial U_i/\partial x_j > 0$,

(ii) $\partial U_i/\partial x_i + \partial U_i/\partial x_j > 0$,

(iii) if $j \neq \emptyset$, then $U_i(C_{i,j}/2, C_{i,j}/2, j) > U_i(0, 0, \emptyset)$.

It implies that (i) all players always prefer to get an additional dollar even if it need be taken from the partner, (ii) they prefer an additional dollar also if the partner gets an additional dollar at the same time, (iii) there is at least one way to share the surplus in any match that both players prefer to staying single (namely to share it equally). That is, players do not forego money because of excessive altruism or wish to waste money because of excessive spite. As result, rational players will share the whole surplus generated within their match and staying single cannot be stable.

The definition of the core for such generalized utilities requires a different formulation than the core for payoff maximizing players, but it remains technically straightforward. On the one hand, the stability of any payoff allocation depends on the matching that applies, since utilities depend on the matching. Hence, the definition of "outcome" need be extended to also include the matching. Define a matching m as a function $m : N \to N \cup \{\emptyset\}$ satisfying, for all $i \in N$, $m(i) \in N_i$ and $m(i) \neq \emptyset \Rightarrow m(m(i)) = i$. Let M be the set of all these matchings. The set of outcomes can now be defined as

$$\begin{aligned}\mathbf{X} = \big\{(\mathbf{x}, m) \in \mathbb{R}_+^N \times M \mid \forall i \in N : m(i) = \emptyset \Rightarrow x_i = 0 \text{ and} \\ \forall w \in W :\ m(w) \neq \emptyset \Rightarrow x_w + x_{m(w)} \leq C_{w,m(w)}\big\}.\end{aligned}$$

On the other hand, when defining stability under generalized utilities, we need to explicitly take into account that it may be preferable to be single than to share the surplus generated by a match in a highly asymmetric way. For, Assumption 4.1 only rules out sharing the surplus equally is preferable to remaining single, but not that every way of sharing the surplus is. To generalize stability correspondingly, let $C_{w,\emptyset} = C_{\emptyset,f} = 0$ for all w, f denote the productivity of single players, and let $U_\emptyset = 0$ denote the utility of the dummy player "$\emptyset$" who represents the partner of an unmatched player.

Definition 4.2 (Core for generalized utility)**.** The GU core is the set of outcomes $(\mathbf{x}|m) \in \mathbf{X}$ such that, for all $i \in N$, all $j \in N_i$, and all $x' \in [0, C_{i,j}]$,

$$U_i(x_i, x_{m(i)}, m(i)) \geq U_i(x', C_{i,j} - x', j) \quad \text{or} \quad U_j(x_j, x_{m(j)}, m(j)) \geq U_j(C_{i,j} - x', x', i).$$

The literature (including Otto and Bolle, 2011) have considered two solution concepts that weaken the stability requirements underlying the core. The *equal-division* *core* (Selten, 1972) relaxes the requirement that if a payoff allocation between two players exists such that both of them benefit by coalescing, then the players will find it. Clearly, finding such payoff allocations is trivial if players maximize payoffs, but for generalized utiltiy, it is not, and hence, weakening the requirement may be reasonable.

Definition 4.3 (Equal-division core)**.** The ED core is the set of outcomes $(\mathbf{x}|m) \in \mathbf{X}$ such that, for all $i \in N$, all $j \in N_i$, and $x' = C_{i,j}/2$,

$$U_i(x_i, x_{m(i)}, m(i)) \geq U_i(x', C_{i,j} - x', j) \quad \text{or} \quad U_j(x_j, x_{m(j)}, m(j)) \geq U_j(C_{i,j} - x', x', i).$$

An even weaker requirement is the "equal split" solution concept, which Otto and Bolle (2011) found to be most descriptive in their analysis. The idea is that players may not try to predict possible payoff allocations in alternative matches at all, arguably due to the uncertainty underlying the necessary negotiations. These players would evaluate the current payoff allocation against their utility if the surplus would be split equally, and they are content if the deviation is not large. We refer to the respective solution set for social preferences as the *equality square*.

Definition 4.4 (Equality square)**.** Fix $\gamma > 0$. The Eq-square is the set of outcomes $(\mathbf{x}|m) \in \mathbf{X}$ such that for all $i \in N$, $U_i(x_i, x_{m(i)}, m(i)) \geq U_i(C_{i,m(i)}/2, C_{i,m(i)}/2, m(i)) - \gamma$.

Note that both, equal-division core and equality square are defined for the generalized utility function. This allows us, in the next section, to also evaluate the combination of social preferences and weaker stability requirements.

5 Evaluation of the core and its variants

The core and its variants yield set "predictions" in the sense that the outcome is predicted to be in a set. For the purpose of actual predictions, this approach is inapplicable, as there will be observations outside the core unless the core includes all observations. On the one hand, this motivates the definition of random utility cores further below, but on the other hand, it implies that we cannot use established approaches such as maximum likelihood or least squares to evaluate the validity of the core. Maximum likelihood cannot be used, as the aggregate likelihood of all variants will be zero, and least squares between say core area and observations cannot be used, as we often observe say B-matching where the core predicts A-matching.[5] A reasonable measure of the distance between A-matching and B-matching is not available. An appropriate measure of the goodness of fit in this context is Selten's score (for original and axiomatic definitions, see Selten, 1972, 1991, and for a critical discussion, see Hey, 1998).

Definition 5.1 (Selten's score)**.** Selten's score of a solution concept is the difference between (i) the relative frequency of observations compatible with the concept and (ii) the share of internally Pareto efficient outcomes compatible with the concept.

Definition 5.2 (Internal Pareto efficiency)**.** An outcome $(\mathbf{x}, m) \in \mathbf{X}$ is internally Pareto efficient if $m(i) \neq \emptyset$ and $x_i + x_{m(i)} = C_{i,m(i)}$ for all $i \in N$.

The model parameters are estimated by maximizing Selten's score jointly over all parameters, using a robust, gradient free algorithm for the initial approach to the maximum, a Newton method to ensure convergence, and various starting values to verify globality of the maximum.

Further, as the concepts discussed here are novel, we follow recent analyses of choice under risk, e.g. Wilcox (2008) and Hey et al. (2010), and determine both their "descriptive validity" and "predictive validity." Thus, we can clarify to which degree possibly superior fit is due to overfitting. Our approach combines cross validation (Burman, 1989; Zhang, 1993) with non-random holdout samples (Keane and Wolpin, 2007). That is, we fit the model to the observations from four of the six treatments, evaluate its fit on the observations from the remaining two treatments, and rotate so that all observations are used in the evaluation stage once. The aggregate Selten score out-of-sample will be our measure of predictive accuracy.

Table 3 lists the Selten scores for six variants of the core, namely all combinations of the three core variants (core, ED core, and Eq-square) as defined above and two utility functions, egoism and linear altruism/spite. The altruistic utility function is defined as $U_i(x_i, x_j, j) = x_i + \alpha x_j + \beta \overline{C}_{ij}$, where α and β are free parameters and $\overline{C}_{ij}$ is the sum of wages (i.e. the productivity) in the other match. In case i is single, i.e. $j = \emptyset$, the utility is $U_i(0,0,\emptyset) = 0 + \beta \cdot \max\{C_{2,1}, C_{2,2}\}$. Table 3b provides the results of

[5] Recall that as defined in Section 2, the matching $\{(W_1, F_1), (W_2, F_2)\}$ is called "A-matching" and $\{(W_1, F_2), (W_2, F_1)\}$ is called "B-matching."

Table 3: Goodness-of-fit measures of the basic structural models

(a) Selten scores (higher is better), descriptively and predictively

	Descriptive validity		Predictive validity	
	Egoism	Altruism	Egoism	Altruism
Eq-Square	0.575	0.612	0.574	0.586
ED Core	0.312	0.512	0.312	0.509
Core	0.122	0.276	0.122	0.253

(b) Non-parametric tests of difference of the Selten scores – out-of-sample

		Model 2					
Model 1	S-Score	Eq-Square, Ego	ED-Core, Ego	Core, Ego	Eq-Square, Altr	ED-Core, Altr	Core, Altr
Eq-Square, Ego	0.57		>>>	>>>	=	=	>>>
ED-Core, Ego	0.31	<<<		>>>	<<<	<<<	>
Core, Ego	0.12	<<<	<<<		<<<	<<<	<<
Eq-Square, Altr	0.59	=	>>>	>>>		>	>>>
ED-Core, Altr	0.51	=	>>>	>>>	<		>>>
Core, Altr	0.25	<<<	<	>>	<<<	<<<	

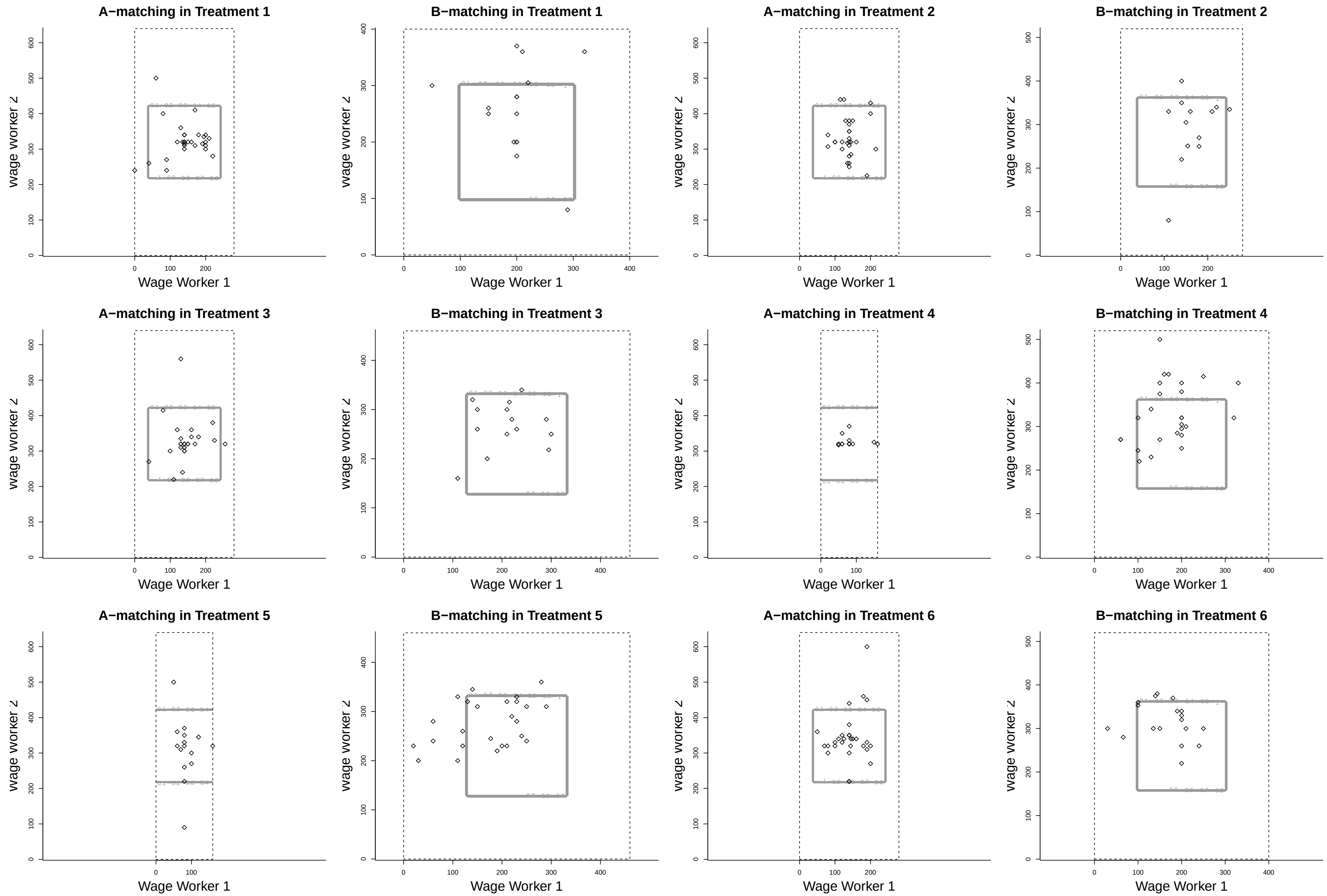
Note: Table 3b report the results of two-sided tests of "model 1" (the row model) against "model 2" (the column model), using Wilcoxon matched-pairs tests of the Selten scores for the 28 sessions. The relation signs $>>>$, $>>$, $>$ indicate that H_0 is rejected at 0.001, 0.02, 0.1 level (respectively), here in favor of model 1, the signs $<<<$, $<<$, $<$ indicate rejection of H_0 in favor of model 2, and $=$ indicates insignificance.

significance tests on the Selten scores. Further analysis confirmed that other functional forms of utility, Fehr-Schmidt inequity aversion and CES utilities, do not improve upon linear altruism in this context (see the supplementary material). The main results concerning model validity (Tables 3 and 3b) an can be summarized as follows.

Result 5.3. *The best fitting models are the equality square and the equal-division core for altruistic preferences. These models are compatible with approximately 80% of the observations but only with 25% of the possible outcomes, and they fit significantly better than all other models (according to Selten's score, at* $\alpha = .01$*).*

In addition, Figure 2 shows that the equality square fits the data adequately, and in particular, in relation to Figure 1 it illustrates the improvement gained on the core. These results suggest that the subjects do not reason as deeply as assumed by the core solution concept. They seem to evaluate outcomes without considering all possible

Figure 2: Contour plots for the equality square (egoistic preferences)



allocations of wages in alternative matches. At most, they seem to consider the alternative match under the simplifying assumption that wages will be split equally, as in the equal division core, but the even simpler equality square seems to provide a more robust explanation at this point (since it works similarly well for both egoistic and altruistic preferences).

However, all of these concepts are "incomplete" in that none of them is compatible with all observations made in the experiment. This raises the question of how the deviations can be explained. Further, the observations compatible with the Eq-square do not seem to be distributed uniformly on it, and the incompatible ones do not seem to be uniformly outside it. Rather, the compatible observations tend to be in the center of the Eq-square, and the incompatible ones tend to be near it. An explanation for both, the occurence of deviations in the first place and this kind of structure of observations, is provided by the random utility concept introduced in the next section.

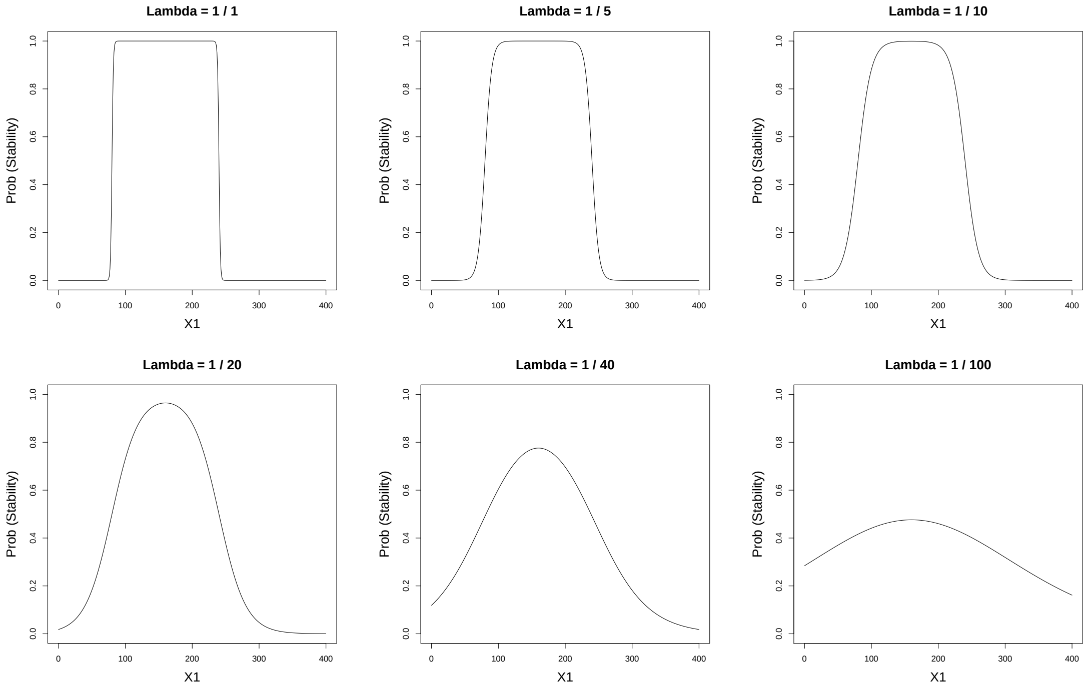
6 Random utility core

Simple bargaining games

Initially, consider a simple bargaining problem. There are two players negotiating the allocation of a cake valued $C > 0$. We define the random utility bargaining problem by adding a random utility component to the outside option. The distribution of the random component will be logistic in our analysis, i.e. the difference of two i.i.d. extreme-value distributed random variables, which corresponds closely with the approach taken in non-cooperative game theory (see e.g. McKelvey and Palfrey, 1995, Goeree and Holt, 1999, Weizsäcker, 2003, Turocy, 2005). Illustrations follow shortly.

Definition 6.1 (Random utility bargaining game)**.** The set of players is $N = \{1,2\}$, the set of possible outcomes is $\mathbf{X} = \{\mathbf{x} \in \mathbb{R}^2_+ \mid x_1 + x_2 \leq C\}$ for some $C > 0$, and the players' disagreement payoffs are $\underline{x}_1, \underline{x}_2 \in [0,C]$ with $\underline{x}_1 + \underline{x}_2 < C$. For both $i \in N$, utilities are $u_i(\mathbf{x}) = x_i$ for all $\mathbf{x} \in \mathbf{X}$ and $\tilde{u}_i(\underline{\mathbf{x}}) = \underline{x}_i + \varepsilon_i$ for the outside option. The distributions of ε_1 and ε_2 are continuous, stochastically independent, and characterized by the cumulative distribution functions F_1 and F_2, respectively.

Figure 3: Stochastic stability in bargaining games for varying precisions λ



Note: The cake size is $C = 400$ and the outside options are $\left(\underline{x}_1, \underline{x}_2\right) = (120, 240)$. The plotted functions are $\Pr(x_1 \geq \underline{x}_1 + \varepsilon_1$ and $x_2 \geq \underline{x}_2 + \varepsilon_2)$, $\varepsilon_1, \varepsilon_2$ being i.i.d. logistic, as functions of x_1 with $x_2 = C - x_1$.

In the unperturbed game, the core $\mathbf{X}^c \subseteq \mathbf{X}$ is the set of individually rational, Pareto efficient allocations.

$$\mathbf{X}^C = \left\{\mathbf{x} \in \mathbf{X} \mid x_i \geq \underline{x}_i \ \wedge \ x_j \geq \underline{x}_j \ \wedge \ x_i + x_j = C\right\}$$

In this case of zero variance, it is appropriate to say that allocation $\mathbf{x} \in \mathbf{X}$ is stable if it is in the core. If utilities are random, however, stability is stochastic. The probability that player i is content with $\mathbf{x}$ is $\Pr(x_i \geq \underline{x}_i + \varepsilon_i)$. For example, if the utility perturbations ε_i have logistic distribution with scale parameter $s = 1/\lambda$, then i is content with probability $1/\left[1 + \exp(\lambda(\underline{x}_i - x_i))\right]$.

Figure 3 plots the probabilities that both players are content with an outcome (x_1, x_2), assuming Pareto efficiency (i.e. $x_2 = C - x_1$) and logistic perturbations. This probability will be called the *stochastic stability* of $\mathbf{x}$,

$$\pi(\mathbf{x}) = \Pr(x_i \geq \underline{x}_i + \varepsilon_i) \cdot \Pr(x_j \geq \underline{x}_j + \varepsilon_j). \tag{1}$$

We say that an outcome $\mathbf{x}$ is *stochastically more stable* than $\mathbf{x}'$ if $\pi(\mathbf{x}) > \pi(\mathbf{x}')$.

This ordering of outcomes generalizes deterministic stability in an intuitive way. As the precision λ tends to infinity, stochastic stability converges pointwise to the stability indicator $\mathbf{1}_{\mathbf{x}\in\mathbf{X}^C}$ of the unperturbed game. The stochastically most stable allocation is generally in the interior of the core of the unperturbed game, and if perturbations are identically distributed for the players, the stochastically most stable outcome is the Nash solution. This is easy to see if perturbations are logistic $F_i(r) = 1/\big(1+\exp(-\lambda_i r)\big)$ for all $r \in \mathbb{R}$ and $i \in N$. In this case, the stochastic stability $\pi(\mathbf{x}) = \big(1+e^{-\lambda_i(x_i-\underline{x}_i)}\big)^{-1} * \big(1+e^{-\lambda_j(1-x_i-\underline{x}_j)}\big)^{-1}$ is maximized if

$$\frac{1+e^{\lambda_i(x_i-\underline{x}_i)}}{1+e^{\lambda_j(1-x_i-\underline{x}_j)}} = \frac{\lambda_i}{\lambda_j}.$$

If errors are i.i.d. logistic, then $\lambda_i = \lambda_j$ and the most stable outcome is the Nash bargaining solution. The following result establishes equivalence between the Nash solution and the stochastically most stable outcome for a general class of distributions.

Lemma 6.2. *Assume the random utility components of all players are i.i.d. with cumulative density F. If F is symmetric, $F(x) = 1-F(-x)$, and has quasi-concave density, then the unique maximizer of $\pi(\mathbf{x})$ is $x_i = (\underline{x}_i + C - \underline{x}_j)/2$ and $x_j = (\underline{x}_j + C - \underline{x}_i)/2$.*

Proof. The first-order condition for $\max_{x_i} F(x_i-\underline{x}_i) * F(C-x_i-\underline{x}_j)$ yields

$$f(x_i-\underline{x}_i)/f(C-x_i-\underline{x}_j) = F(x_i-\underline{x}_i)/F(C-x_i-\underline{x}_j).$$

The claimed solution implies $x_i - \underline{x}_i = C - x_i - \underline{x}_j =: x'$ and hence satisfies the condition. Next, $\underline{x}_i + \underline{x}_j < C$ implies $x' = (C-\underline{x}_i-\underline{x}_j)/2 > 0$. The second-order condition (for the claimed solution to be a maximum) is $2f'(x')F(x') < 2f(x')f(x')$. By symmetry and quasi-concavity, $x' > 0$ implies $f'(x') \leq 0$; since all other terms are positive, the condition holds. Finally, consider the case $x_i - \underline{x}_i \neq x'$, and without loss, assume $x_i - \underline{x}_i > x'$. Hence, $C - x_i - \underline{x}_j < x'$. By symmetry and quasi-concavity, $f(x_i-\underline{x}_i) \leq f(C-x_i-\underline{x}_j)$, and by monotonicity, $F(x_i-\underline{x}_i) > F(C-x_i-\underline{x}_j)$; hence, the first-order condition is violated, which proves uniqueness. □

Further, outcomes close to the Nash solution are stochastically more stable than distant outcomes, outcomes in the core are stochastically more stable than outcomes outside of it, and outcomes close to the core are stochastically more stable than outcomes distant to the core. As these are exactly the characteristics of the data described above that we wish to capture, we will now define a solution concept where the probability that outcome $\mathbf{x}$ results is proportional to its stochastic stability. Specifically, the *random utility core* is the probability density $f_C \in \Delta\big(\mathrm{PF}(\mathbf{X})\big)$ on the Pareto frontier that is proportional to the above measure of stochastic stability.

$$f_C(\mathbf{x}) = \pi(\mathbf{x}) / \int_{\mathrm{PF}(\mathbf{X})} \pi\big(\tilde{\mathbf{x}}\big) d\tilde{\mathbf{x}} \tag{2}$$

As stated, the integration is along the Pareto frontier defined as

$$\mathrm{PF}(\mathbf{X}) = \{\mathbf{x} \in \mathbf{X} \mid \mathbf{x} \geq \mathbf{x}' \ \forall \mathbf{x}' \in \mathbf{X}\}. \tag{3}$$

Obviously, Pareto efficiency could be relaxed as well, similar to the way individual rationality has been relaxed. We have not done so here, because allocating the whole cake does not seem to be an issue in bargaining. The subjects in our experiment manage to solve this computationally simple task in almost all cases. The deviations from the core are therefore due to some other form of noise, arguably due to randomness of utility as defined above. The following establishes a simple axiomatic foundation of the random utility core.

Proposition 6.3. *For bargaining games (Def. 6.1), the following statements are equivalent.*

1. *f_C satisfies Eq.* (2) *for* $\pi(\mathbf{x}) = F_1(x_1 - \underline{x}_1) * F_2(x_2 - \underline{x}_2)$.
2. *f_C satisfies the following conditions.*
   - *A*1 *Continuity and Pareto efficiency: f_C is the density of a continuous distribution on PF*$(\mathbf{X})$.
   - *A*2 *Proportional stability: $f_C(\mathbf{x})$ is proportional to the probability that all players prefer $\mathbf{x}$ to their outside option, i.e. to* $\Pr\big(u_i(\mathbf{x}) \geq \tilde{u}_i(\underline{\mathbf{x}}) \ \forall i\big)$.

Proof. 2. $\Rightarrow$ 1.: By the definition of the game, ε_1 and ε_2 are independent, and thus $\Pr\big(u_i(\mathbf{x}) \geq \tilde{u}_i(\underline{\mathbf{x}}) \ \forall i\big) = \Pr\big(u_1(\mathbf{x}) \geq \tilde{u}_1(\underline{\mathbf{x}})\big) \cdot \Pr\big(u_2(\mathbf{x}) \geq \tilde{u}_2(\underline{\mathbf{x}})\big) =: \pi(\mathbf{x})$. A2 implies

$f_C(\mathbf{x}) = a \cdot \pi(\mathbf{x})$ for some $a > 0$. Finally, since f_C is a density with support only on the Pareto frontier (A1), $a = 1/\int_{\mathrm{PF}(\mathbf{X})} \pi(\mathbf{x})d\mathbf{x}$. 1. $\Rightarrow$ 2. can be verified easily. $\square$

Note that A2 implies independence of irrelevant alternatives (IIA), i.e. $f_C(\mathbf{x}'|\mathbf{X}') \cdot f_C(\mathbf{x}|\mathbf{X}'') = f_C(\mathbf{x}'|\mathbf{X}'') \cdot f_C(\mathbf{x}|\mathbf{X}')$ for all $\mathbf{x}, \mathbf{x}' \in \mathbf{X}'$ and all measurable $\mathbf{X}' \subseteq \mathbf{X}'' \subseteq \mathrm{PF}(\mathbf{X})$.

Assignment games

Assignment games generalize bargaining games by endogenizing the outside options. For every player, each partner other than his current one represents an outside option, while the values of these outside options depend on the payoff allocations negotiated in their matches. The more my prospective partners in their current matches make, the less the outside options are worth to me (for further discussion, see e.g. Otto and Bolle, 2011). Aside from taking these changes into account, the above definition of the random utility core generalizes immediately. In particular, we maintain the assumption that the utilities of the outside options are random (e.g. logistic in the logit core). The notation on assignment games introduced in Section 4 is maintained.

Definition 6.4 (Random utility assignment game)**.** For each outcome $(\mathbf{x}, m) \in \mathbf{X}$, the utility of $i \in N$ is $u_i(\mathbf{x}, m) = U_i(x_i, x_{m(i)}, m(i))$. The utility of the blocking coalition (i, j) with wages $(x_i, x_j) \in \mathbb{R}^2_+$ is $\tilde{u}_i(x_i, j) = U_i(x_i, x_j, j) + \varepsilon_{i,j}$, and the utility of the outside option is $u_i(x_i, \emptyset) = U_i(0, 0, \emptyset) + \varepsilon_{i,\emptyset}$. The distributions of $\varepsilon_{i,j}$ are continuouos and stochastically independent over all $i \in N$ and $j \in N_i$.

As before, we define stochastic stability as the probability that the allocation is stable, i.e. that no pair of players can rematch profitably. In the random utility assignment game, the stochastic stability of outcome $(\mathbf{x}, m)$ therefore is

$$\pi(\mathbf{x}, m) = \Pr\big(\forall i \in N, \forall j \in N_i, \forall x' \in [0, C_{i,j}] : \\ u_i(\mathbf{x}, m) \geq \tilde{u}_i(x', j) \text{ or } u_j(\mathbf{x}, m) \geq \tilde{u}_j(C_{i,j} - x', i)\big), \quad (4)$$

and exploiting all independence assumptions in Def. 6.4, it simplifies to

$$\pi(\mathbf{x},m) = \prod_{i\in N}\prod_{j\in N_i}\int_{\mathbb{R}}\int_{\mathbb{R}} f_{i,j}(\varepsilon_{i,j})f_{j,i}(\varepsilon_{j,i})\ \mathbf{1}\big\{\forall x' \in [0, C_{i,j}] : \\ U_i(x_i, x_{m(i)}, m(i)) \geq U_i(x', C_{i,j} - x', j) + \varepsilon_{i,j} \text{ or} \\ U_i(x_j, x_{m(j)}, m(j)) \geq U_j(C_{i,j} - x', x', i) + \varepsilon_{j,i}\big\} d\varepsilon_{i,j} d\varepsilon_{j,i}. \quad (5)$$

Intuitively, at the beginning of the assignment game, the utility perturbations $(\varepsilon_{i,j})$ are drawn. One may think of $\varepsilon_{i,j}$ as a measure of the "chemistry" between i and j (in the eyes of i). The players then play the assignment game and evaluate possible allocations by comparing them to the outside options plus the utility perturbations. The stochastic stability is the ex-ante probability that a given allocation will be stable.

The random utility core is again defined as the density of continuous distribution on the Pareto frontier. In general, we assume that players will be able match completely and to allocate the whole productivity of their respective match,[6] i.e. that outcomes satisfy internal Pareto efficiency as defined in Def. 5.2.

Define IPF($\mathbf{X}'$) as the set of internally Pareto efficient allocations in $\mathbf{X}' \subseteq \mathbf{X}$, and $M^* \subset M$ as the set of complete matchings. Thus, using $\mathbf{X}_m = \{\mathbf{x} \in \mathbb{R}^n \mid (\mathbf{x}, m) \in \mathbf{X}\}$, the random utility core is defined as

$$f_C(\mathbf{x},m) = \pi(\mathbf{x},m) / \sum_{m\in M^*}\int_{\mathrm{IPF}(\mathbf{X}_m)} \pi\big(\tilde{\mathbf{x}},m\big) d\tilde{\mathbf{x}}. \quad (6)$$

Clearly, this definition implies that the random utility core converges pointwise to the uniform distribution on the core if the utility variances approach zero. The assumption that stochastic stability and outcome density are proportional is particularly simple and it turns out to fit our data well (see below). Alternative assumptions may prove appropriate in alternative classes of games.

Proposition 6.5. *For any assignment game, the following statements are equivalent.*

1. *f_C is the random utility core defined in Eqs.* (4) *and* (6).

[6]These assumptions can be relaxed, of course, but they reflect the observations made in Section 3. In turn, we do not assume external Pareto efficiency, which would be akin to the assumption of social efficiency and seems too strong in light of the above observations.

2. *f_C satisfies the following conditions.*

   A1 Continuity and internal Pareto efficiency: f_C is the density of a continuous distribution on the set outcomes satisfying internal Pareto efficiency.

   A2 Proportional stability: $f_C(\mathbf{x},m)$ is proportional to the stochastic stability $\pi(\mathbf{x},m)$.

Proof. The proof is very similar to that of Prop. 6.3 and therefore skipped. □

Finally, we provide the stochastic stabilities of the core variants discussed before. The equal-division core for random utility induce the stochastic stability

$$\pi(\mathbf{x},m) = \Pr\big(\forall i \in N, \forall j \in N_i, x' = C_{i,j}/2 : \\ \tilde{u}_i(x',j) \leq u_i(\mathbf{x},m) \text{ or } \tilde{u}_j(C_{i,j}-x',j) \leq u_j(\mathbf{x},m)\big), \quad (7)$$

in conjunction with Eq. (4). The equality square obtains if the stability is defined as

$$\pi(\mathbf{x},m) = \Pr\big(\forall i \in N: \; u_i(\mathbf{x},m) > \tilde{u}_i(C_{i,m(i)}/2, m(i))\big). \quad (8)$$

7 Evaluation of the random utility core

The section describes and discusses the estimation results for the random utility cores with logistic errors, i.e. for the logit core and its (logit) variants. We contrast the logit cores with two similarly intuitive but technically simpler econometric models of the negotiated matches and wages. These will be regression models, which explain the distribution of matches and wages without explicit references to utility functions, and random behavior models, which explain deviations from the core as the result of trembles. The maximization procedure is similar to above, i.e. the likelihood is maximized jointly over all parameters by first gradient-free and second Newton methods, and a variety of starting values is used to verify globability of the maximum. As before, we determine both descriptive accuracy and predictive accuracy, in order to verify the validity of the best-fitting models. Table 4 lists absolute values of the log-likelihoods for all models, Table 5 describes the results of likelihood-ratio tests (nested and non-nested Vuong tests), and Figure 4 provides the contour plots of the best-fitting model. All parameter estimates are provided as supplementary material.

Descriptive and predictive accuracy

Initially, we focus on the goodness-of-fit of the random utility cores, see Tables 4a and 5a. The main results can be summarized as follows.

Result 7.1. *As for the random utility concepts, the equality square for egoistic preferences and the equal-division core for altruistic preferences do not differ significantly ($p > 0.1$) and the ED-core fits significantly better than all other concepts ($p < 0.1$). Further, the difference between descriptive and predictive accuracy is insignificant for these concepts.*

That is, the core concept is rejected. Subjects do not consider the whole range of payoffs from alternative matches when they evaluate their current situation. Based on the above results, we cannot conclusively discriminate between two forms of weaking the core concept, however. The equality square and the equal division core for altruistic preferences seem to fit the data similarly well. At first glance, this is surprising, as the experiment had been explicitly designed to distinguish between these possibilities. The main design principle was that the players can always be segragated into "strong" and "weak" players based on payoffs from the outside option (the alternative match). This principle implies that if the alternative match is strategically relevant, then the equality square cannot fit well. In turn, if wage equality actually is important, then the payoffs from the alternative matches cannot be found to be strategically relevant. The fact that both classes of models are of similar econometric validity therefore indicates that indeed both kinds of motives affect "stability" of outcomes.

This hypothesis is analyzed next, by considering a generalized model nesting both of these alternatives. We refer to this generalized concept as the *Eq-ED core*. Its stability is the weighted mean of the stochastic stabilities π_{Eq} in the Eq-core for altruistic preferences and π_{ED} in the ED-core for egoistic preferences (which are the best-fitting

Table 4: Goodness of fit: Overview of all models

(a) Goodness-of-fit $|LL|$ of the random utility models

	Descriptive validity		Predictive validity	
	Egoism	Altruism	Egoism	Altruism
Eq-Logit	3003.89	2987.35	3007.54	3013.29
ED-Logit	3156.63	2979.29	3157.7	2992.64
Logit	3157.26	2998.09	3160.67	3013.57
Eq-ED-Logit	2982.87	2945.37	2986.07	2956.98

(b) Parameter estimates of the Eq-ED-logit model with altruism

λ_{ED}	α	β	λ_{Eq}	μ	LL
32.8933	−0.489	−0.2537	13.5567	0.0294	−2945.37
(3.6258)	(0.053)	(0.0537)	(1.4161)	(0.0084)	

Note: λ_{ED} and λ_{Eq} are the precision parameters of the ED and Eq components (resp.), α, β are the altruism coefficients, and μ is the mixture weight as defined in Eq. (9)

(c) Goodness-of-fit $|LL|$ of the random behavior models

	Descriptive validity		Predictive validity	
	Egoism	Altruism	Egoism	Altruism
Eq-RBehav	3085.07	3057.39	3085.70	3068.21
ED-RBehav	3225.18	3106.25	3225.79	3330.29
RBehav	3253.32	3191.34	3255.17	3208.04

(d) Goodness-of-fit of the regression models

	Descriptive $|LL|$	Predictive $|LL|$
Regression on treatment parameters		
AtheorRegr	2971.60	3052.92
AtheorRegr2	2964.04	3038.54
Regression on theoretically relevant parameters		
RedForm	3006.40	3015.56
RedForm2	2979.82	3120.54

concepts out of sample), with weights $\mu \in [0,1]$.

$$\pi_{Eq-ED}(\mathbf{x},m) = \mu \cdot \pi_{ED}(\mathbf{x},m) + (1-\mu) \cdot \pi_{Eq}(\mathbf{x},m) \tag{9}$$

The equal-division core obtains for $\mu = 1$, the equality square obtains for $\mu = 0$, and other values of μ yield mixtures of these concepts. Note that such models merging notions of stability cannot be defined for the unperturbed core, as stability is binary there, and a mixed stability of say 0.3 is meaningless in this context. Such mixtures can be analyzed only using concepts based on the notion of stochastic stability such as the one defined in Eq. (6). The parameter estimates are provided in Table 4b, the log-likelihoods both descriptively and predictively are provided in Table 4a, and the results of the corresponding likelihood-ratio tests are provided in Table 5.

The estimated mixture weight is $\mu = 0.029$, which is close to but significantly different ($p < .01$) from zero. Figure 4 shows how the secondary influence of equal division in alternative matches affects predictions in the estimated model. It broadens the predicted range of outcomes into the direction predicted by the core (i.e. equal-division core). As a result, the log-likelihood improves highly significantly (in both nested and non-nested likelihood-ratio tests) while overfitting of the generalized model is insignificant (in non-nested likelihood ratio tests). Therefore, the following result.

Result 7.2. *The subjects' main criterion for stability is equality of incomes within matches. The potential payoff from the alternative match is of secondary, but significant relevance. The respective model, the Eq-ED core fits significantly better than all other models and it does not overfit significantly.*

To conclude this part of the analysis, let us note that the estimated utility parameters are $\alpha = -0.489$ and $\beta = -0.2537$ (see Table 4b). The degree of altruism actually is negative, which implies competitive and aggressive bargaining in 2×2 assignment games. We consider this to be reasonable in the present context, even though altruism seems to be positive in other contexts such as dictator games.

Table 5: Results of likelihood ratio tests

(a) Comparison of the basic structural models with egoism and altruism – out-of-sample

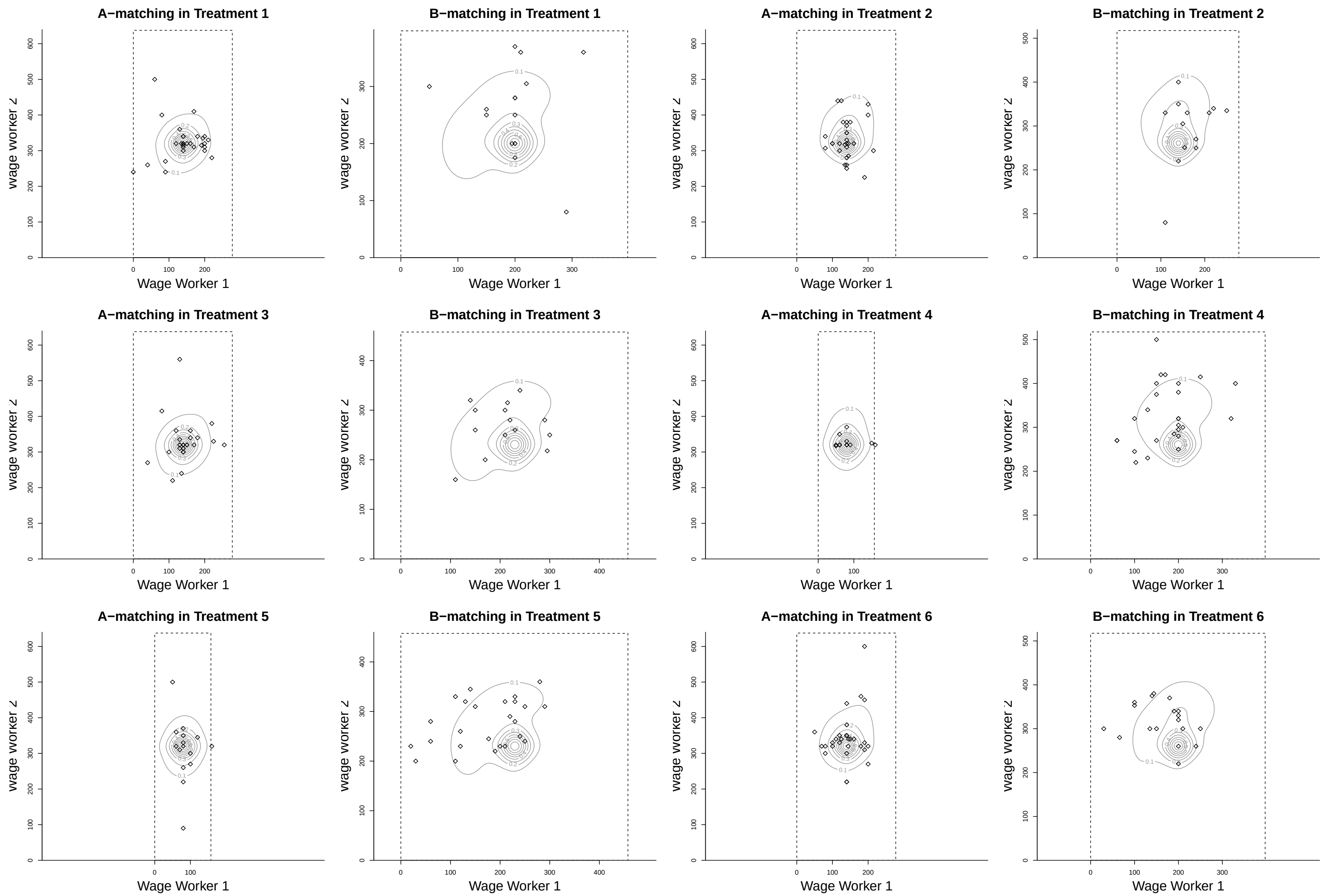
Model 1	LL	Model 2									
		Eq-Logit, Ego	ED-Logit, Ego	Logit, Ego	RBehav, Ego	Eq-ED-Logit, Ego	Eq-Logit, Altr	ED-Logit, Altr	Logit, Altr	RBehav, Altr	Eq-ED-Logit, Altr
Eq-Logit, Ego	−3007.54		>>>	>>>	>>>	<	=	=	=	>>>	<<<
ED-Logit, Ego	−3157.7	<<<		=	>>>	<<<	<<<	<<<	<<<	>>	<<<
Logit, Ego	−3160.67	<<<	=		>>>	<<<	<<<	<<<	<<<	>	<<<
RBehav, Ego	−3255.17	<<<	<<<	<<<		<<<	<<<	<<<	<<<	<	<<<
Eq-ED-Logit, Ego	−2986.07	>	>>>	>>>	>>>		>>	=	>>	>>>	<<
Eq-Logit, Altr	−3013.29	=	>>>	>>>	>>>	<<		<	=	>>>	<<<
ED-Logit, Altr	−2992.64	=	>>>	>>>	>>>	=	>		>	>>>	<<<
Logit, Altr	−3013.57	=	>>>	>>>	>>>	<<	=	<		>>>	<<<
RBehav, Altr	−3208.04	<<<	<<	<	>	<<<	<<<	<<<	<<<		<<<
Eq-ED-Logit, Altr	−2956.98	>>>	>>>	>>>	>>>	>>	>>>	>>>	>>>	>>>	

(b) Comparison with the alternative models – out-of-sample

Model 1	LL	Model 2									
		Eq-Logit, Altr	ED-Logit, Altr	Logit, Altr	Atheor Regr	Atheor Regr2	Reduced Form	Reduced Form2	Eq-RBehav, Altr	RBehav, Altr	Eq-ED-Logit, Altr
Eq-Logit, Altr	−3013.29		<	=	>	=	=	>>>	>>	>>>	<<<
ED-Logit, Altr	−2992.64	>		>	>>>	>>	>	>>>	>>>	>>>	<<<
Logit, Altr	−3013.57	=	<		>	=	=	>>>	>>>	>>>	<<<
Atheor Regr	−3052.92	<	<<<	<		<<	<<	>>	=	>>>	<<<
Atheor Regr2	−3038.54	=	<<	=	>>		<	>>>	=	>>>	<<<
Reduced Form	−3015.56	=	<	=	>>	>		>>>	>	>>>	<<
Reduced Form2	−3120.54	<<<	<<<	<<<	<<	<<<	<<<		<	>>	<<<
Eq-RBehav, Altr	−3068.21	<<	<<<	<<<	=	=	<	>		>>>	<<<
RBehav, Altr	−3208.04	<<<	<<<	<<<	<<<	<<<	<<<	<<	<<<		<<<
Eq-ED-Logit, Altr	−2956.98	>>>	>>>	>>>	>>>	>>>	>>	>>>	>>>	>>>	

Note: The notation is the same as in Table 3. The results are obtained by non-parametric (i.e. distribution-free) implementations of the Vuong (1989) test for nested/non-nested models (as appropriate). That is, with $(f_s)_{s=1}^{28}$ and $(g_s)_{s=1}^{28}$ denoting the log-likelihoods of two competing models in the 28 independent sessions, we evaluate the null hypothesis $H_0 : \ln(f_i/g_i) = 0$ using Wilcoxon signed-rank tests.

Figure 4: Contour plots of the relative stochastic stabilities for the Eq-ED logit core with altruistic preferences. The iso-lines connect outcomes with the same stochastic stability and hence the same predicted density. Outcomes along the the outmost line (at “0.1”) have 10% of the stochastic stability (and density) of the stochastically most stable outcome in the respective game.



Comparison with alternative models

Next, we compare the random utility cores with alternative approaches of modeling wages and matches in 2×2 assignment games. The purpose is to examine the degree to which random utility cores improve upon alternative, possibly simpler concepts. The first alternative is the random behavior core, which is based on the idea that deviations occur with positive probability but are not further structured (i.e. they are uniform).

Definition 7.3 (Random behavior core)**.** Let $\pi^* : \mathbf{X} \to \{0,1\}$ be a binary stability indicator on the set of outcomes (such as those implied by the GU core, Def. 4.2, or ED core, Def. 4.3). The random behavior core is the density of a continuous distribution on the set outcomes satisfying internal Pareto efficiency satisfying, for some $\varepsilon > 0$,

$$f_C(\mathbf{x},m) = \begin{cases} 1/(1+\varepsilon), & \text{if } \pi^*(\mathbf{x},m) = 1, \\ \varepsilon/(1+\varepsilon), & \text{if } \pi^*(\mathbf{x},m) = 0. \end{cases} \tag{10}$$

As ε tends to zero, the random behavior core converges uniformly to the uniform distribution on the core represented by the underlying stability indicator. The technical difference to the random utility core are that the density is an affine transformation of the *unperturbed* stability, and hence it is constant inside the core, on the one hand, and constant outside of it, on the other. According to the random utility core, on the contrary, outcomes in the center of the core are "more stable" than those on its boundary, and outcomes close to the core are more stable than distant ones. Notions of random behavior and in particular of trembles are popular in analyses of choice under risk, see e.g. Conte et al. (2011). More general discussions can be found in Hey (2005) and Loomes (2005). Comparative analyses of random behavior and random utility in non-cooperative games (e.g. McKelvey and Palfrey, 1995, and Weizsäcker, 2003) suggest that utility is random, not behavior.

We examine the random behavior models of the core in all of its six variants considered before. Table 4c provides the goodness-of-fit measures, Table 5b provides the results of the corresponding likelihood-ratio tests, and the parameter estimates are provided as supplementary material. The main result is that random behavior fits worse than random utility in the 2×2 assignment games.

Result 7.4. *All random behavior models fit significantly worse ($p < .01$) than their random utility counterparts, both descriptively and predictively.*

Thus, the notion that utility is random, which explains both structure of observations inside the core and deviations from the core, is more accurate than modeling deviations from the core as random trembles.

In order to further underline that random utility modeling is indeed appropriate in this context, let us finally consider the best known alternative: regression modeling. Clearly, regression analyses feature highly in experimental economics, seemingly because they are computationally less intricate than structural analyses. However, Keane (2010) and Rust (2010), amongst others, argue that econometric models that do not capture the strategic aspects of the interaction in question risk overfitting and hence lack robustness. We investigate this hypothesis on four regression models.

In all cases, the dependent variables are the binomial variable matching $M \in \{A, B\}$ and the worker wages w_1, w_2. Models *AtheorRegr* and *AtheorRegr2* are atheoretical models regressing the observations (M, w_1, w_2) on the treatment variables $C_{1,1}$, $C_{1,2}$, $C_{2,1}$ (recall that $C_{2,2}$ is held constant in all treatments). Matching is modeled by logistic regression, and wages are modeled as continuous variables with truncated normal errors (since the support of the wages is truncated).

$$\Pr(A) = 1/\big(1+\exp(-I_0 - p_{0,C_{1,1}}C_{1,1} - p_{0,C_{1,2}}C_{1,2} - p_{0,C_{2,1}}C_{2,1})\big) \tag{11}$$

$$f(w_1) = \Pr(A)\cdot f(w_{1|A}) + (1-\Pr(A))\cdot f(w_{1|B}) \tag{12}$$

$$f(w_2) = \Pr(A)\cdot f(w_{2|A}) + (1-\Pr(A))\cdot f(w_{2|B}) \tag{13}$$

with

$$w_{i|M} \sim \mathcal{N}\big(I_{i|M} + p_{i|M,C_{1,1}}\, C_{1,1} + p_{i|M,C_{1,2}}\, C_{1,2} + p_{i|M,C_{2,1}}\, C_{2,1}, \sigma^2_{i|M}\big) \quad \text{with support } \big[0, \overline{w}_{i|M}\big]$$

for all $i \in \{1,2\}$ and $M \in \{A,B\}$, using $(\overline{w}_{1|A}, \overline{w}_{2|A}, \overline{w}_{1|B}, \overline{w}_{2|B}) = (C_{1,1}, C_{2,2}, C_{1,2}, C_{2,1})$. In the model *AtheorRegr*, the wage coefficients are independent of the resulting match, i.e. $p_{1|A} = p_{1|B}$ and $p_{2|A} = p_{2|B}$, which yields 14 parameters overall. In the model *AtheorRegr2*, this restriction is lifted, which yields 24 parameters overall. As before, parameters are estimated by maximizing the likelihood jointly over all parameters.

Table 4d presents the results. The difference between descriptive and predictive log-likelihood exceeds 50 points for both models. Although regression on treatment parameters is commonly practiced in the literature, this discrepancy renders the approach inappropriate in our case.

Alternatively, we consider two regression models representing the best fitting model from above, Eq-ED core, in reduced form. These models are labeled *RedForm* and *RedForm2*, and in relation to the Eq-ED logit core, they assume normal rather than logistic errors and egoistic preferences (introducing the notion altruism requires a structural approach). The wages are modeled as above, Eqs. (12) and (13), but now $w_{i|M}$ has the equal split as the intercept and the difference in the partners' outside options as secondary influence. The matching probabilities are functions of the relative social efficiencies of the matches and of the asymmetry indicator $C_{1,2} = C_{2,1}$ (as discussed in the experimental design). Formally, the models are defined as follows.

$$\Pr(A) = 1/\big(1+\exp(-I_0 - p_{0,\mathrm{Eff}}\ (C_{1,1}+C_{2,2}-C_{1,2}-C_{2,1}) - p_{0,\mathrm{Diff}}\ \mathrm{I}_{C_{1,2}=C_{2,1}}\big)$$

with $f(w_1)$ and $f(w_1)$ as in Eqs. (12) and (13), now using

$$\begin{aligned}
w_{1|A} &\sim \mathcal{N}\big(C_{1,1}/2 + p_{1|A}\ (C_{2,1}-C_{1,2})/2, \sigma^2_{1|A}\big) && \text{with support } [0, C_{1,1}] \\
w_{2|A} &\sim \mathcal{N}\big(C_{2,2}/2 + p_{2|A}\ (C_{1,2}-C_{2,1})/2, \sigma^2_{2|A}\big) && \text{with support } [0, C_{2,2}] \\
w_{1|B} &\sim \mathcal{N}\big(C_{1,2}/2 + p_{1|B}\ (C_{2,2}-C_{1,1})/2, \sigma^2_{1|B}\big) && \text{with support } [0, C_{1,2}] \\
w_{2|B} &\sim \mathcal{N}\big(C_{2,1}/2 + p_{2|B}\ (C_{1,1}-C_{2,2})/2, \sigma^2_{2|B}\big) && \text{with support } [0, C_{2,1}].
\end{aligned}$$

Model *RedForm* invokes $p_{1|A} = p_{1|B} =: p_1$ and $p_{2|A} = p_{2|B} =: p_2$, which yields seven parameters, while *RedForm2* allows for asymmetry between A and B matches and has eleven parameters. The goodness-of-fit measures for all models are listed in Table 4d. Most interestingly, the descriptive validities of *RedForm* and *RedForm2* are comparable to those of the Eq logit square and the Eq-ED logit square, respectively, for egoistic preferences—i.e. they capture the theoretical intuition they were supposed to represent in reduced form. However, they do not capture it explicitly, and as can be seen from their poorer predictive fit (in particular of *RedForm2*), they do so less robustly. We conclude that both kinds of regression models are less appropriate than the core.

Result 7.5. *Three of the four regression models improve upon the logit equality square in-sample, but all of them overfit drastically and their predictive accuracies are comparably poor (they are significantly worse than those of all random utility models allowing for altruistic preferences). The fourth regression model largely prevents overfitting, but both its desriptive and its predictive accuracy are significantly worse than that of the Eq-ED core.*

Absolute fit of random utility cores

The last issue to be analyzed is the absolute level of the goodness of fit. Does the (Eq-ED) logit core capture the data quantitatively? The short answer is: Yes, it does. First, consider the Cox-Snell pseudo-R^2 measure of the goodness of fit,

$$\tilde{R}^2 = 1 - \left(\frac{L(\text{Baseline})}{L(\text{Model})}\right)^{2/N},$$

with L being the likelihood function. The Baseline model is the model predicting uniform randomization in our case, which has the log-likelihood -3276.27, and the number of observations is $N = 257$. The resulting value for the Eq-ED model (with altruism) is $\tilde{R}^2 = 0.9239$, which indicates a good absolute fit.

This can be tested for directly using chi-square and G tests of goodness of fit. To this end, we measure the distance of observations from the equal split, and then test if the observed relative frequencies are as predicted for all classes of distances. This allows us to test if there are systematic biases for some or all ranges of distance. Let O denote the set of observations, and for all $o \in O$, let $d(o)$ denote the Euclidean distance of the negotiated wages to the equal split under the respective matching. To apply chi-square and G tests, the distances have to be classified, using class widths such that most classes are well populated. In our context, an appropriate number of classes is six with a class width of 23. That is, observation o is in the class $c(o) = 1$ if $d(o) \leq 23$, $c(o) = 2$ if $23 < d(o) \leq 46$, and so on, truncated above at six. The class width was chosen such that focal distances such as 50 or 100 are not integral multiples of it. We verified that the test results are robust to varying both class width and number of classes (subject to the feasibility constraints of chi-square tests). The main result is as follows.

Result 7.6. *The best fitting random utility model (Eq-ED core for altruism) does not differ significantly from the observations based on chi-square and G tests of goodness of fit (the p-values are .055 and .064, respectively). All other models are rejected highly significantly ($p < .01$).*

To summarize, the random utility model Eq-ED core is an appropriate explanation of the data, as it actually fits (Result 7.6), it is a robust, not overfitted explanation, as its predictive accuracy does not differ significantly from its predictive accuracy (Result 7.2), and it fits significantly better than all other models (Results 7.4 and 7.5).

8 Conclusion

This paper has introduced the random utility core as a positive solution concept for cooperative games and evaluated its validity based on experimental data from 2×2 assignment games. We have seen that the random utility core fits significantly better than random behavior models and regression models. Since it also fits absolutely well, Figure 4 and Result 7.6, we conclude that the notion of random utility, which is well-established in experimental analyses on decision theory and non-cooperative game theory, is applicable to cooperative games, as well.

The best fitting model is a simplified variant of the core. It reduces the computational burden imposed on the subjects, which relates to the validity of level-k models found in analyses of non-cooperative games (Stahl and Wilson, 1995; Ho et al., 1998; Bosch-Domenech et al., 2002; Costa-Gomes and Crawford, 2006). Thus, we have shown that much of the recent advances in positive modeling of non-cooperative games—besides random utility and level-k theory, we also modeled social preferences—are equally helpful to obtain predictive models for games in the realm of cooperative theory. Thus, structure of preferences and level of reasoning in cooperative games seems to be rather similar to that in non-cooperative games, although the solution concepts employed are fundamentally different.

Since random utility is a novel approach in modeling cooperative games, there is ample opportunity for further research. In addition to analyses of the random utility core in bargaining games, further research may also define values of games with random utility, and it may explore the link between stochastic stability and outcome density both empirically and theoretically (possibly deriving such links from non-cooperative implementations). Perhaps most interestingly, further research may evaluate the validities of predictions obtained from cooperative and non-cooperative models in comparative studies, in order to map out their respective fields of application and to define new concepts modeling the key insights from both branches. This seems to be promising, as many real-world negotiations are less structured than non-cooperative bargaining games, and hence may be modeled more accurately using cooperative games.

References

Bosch-Domenech, A., Montalvo, J., Nagel, R., and Satorra, A. (2002). One, two,(three), infinity,...: Newspaper and lab beauty-contest experiments. *American Economic Review*, 92(5):1687–1701.

Burman, P. (1989). A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods. *Biometrika*, 76(3):503.

Camerer, C. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton, NJ: Princeton University Press.

Camerer, C., Ho, T., and Chong, J. (2004). A cognitive hierarchy model of games. *Quarterly Journal of Economics*, 119(3):861–898.

Chen, Y. and Sönmez, T. (2002). Improving efficiency of on-campus housing: An experimental study. *The American Economic Review*, 92(5):1669–1686.

Chen, Y. and Sönmez, T. (2006). School choice: an experimental study. *Journal of Economic Theory*, 127(1):202–231.

Conte, A., Hey, J., and Moffatt, P. (2011). Mixture models of choice under risk. *Journal of Econometrics*, 162(1):79–88.

Costa-Gomes, M. and Crawford, V. (2006). Cognition and behavior in two-person guessing games: An experimental study. *American Economic Review*, 96(5):1737–1768.

Costa-Gomes, M., Crawford, V., and Iriberri, N. (2009). Comparing models of strategic thinking in van huyck, battalio, and beil's coordination games. *Journal of the European Economic Association*, 7(2-3):365–376.

Crawford, V. and Knoer, E. (1981). Job matching with heterogeneous firms and workers. *Econometrica*, 49(2):437–50.

Driessen, T. (1998). A note on the inclusion of the kernel in the core of the bilateral assignment game. *International Journal of Game Theory*, 27(2):301–303.

Gale, D. and Shapley, L. (1962). College admissions and the stability of marriage. *American Mathematical Monthly*, 69(1):9–15.

Goeree, J. and Holt, C. (1999). Stochastic game theory: for playing games, not just for doing theory. *Proc Natl Acad Sci US A*, 96(19):10564–10567.

Goeree, J. and Holt, C. (2004). A model of noisy introspection. *Games and Economic Behavior*, 46(2):365–382.

Halaburda, H. (2010). Unravelling in two-sided matching markets and similarity of preferences. *Games and Economic Behavior*, 69(2):365–393.

Hamers, H., Klijn, F., Solymosi, T., Tijs, S., and Pere Villar, J. (2002). Assignment games satisfy the coma-property. *Games and Economic Behavior*, 38(2):231–239.

Hey, J. (1998). An application of Selten's measure of predictive success. *Mathematical Social Sciences*, 35(1):1–15.

Hey, J. (2005). Why we should not be silent about noise. *Experimental Economics*, 8(4):325–345.

Hey, J., Lotito, G., and Maffioletti, A. (2010). The descriptive and predictive adequacy of theories of decision making under uncertainty/ambiguity. *Journal of risk and uncertainty*, 41(2):81–111.

Ho, T., Camerer, C., and Weigelt, K. (1998). Iterated dominance and iterated best response in experimental "*p*-beauty contests". *American Economic Review*, 88(4):947–969.

Kagel, J. and Roth, A. (2000). The dynamics of reorganization in matching markets: A laboratory experiment motivated by a natural experiments. *Quarterly Journal of Economics*, 115(1):201–235.

Keane, M. (2010). A structural perspective on the experimentalist school. *The Journal of Economic Perspectives*, 24(2):47–58.

Keane, M. and Wolpin, K. (2007). Exploring the usefulness of a nonrandom holdout sample for model validation: Welfare effects on female behavior. *International Economic Review*, 48(4):1351–1378.

Kelso Jr, A. and Crawford, V. (1982). Job matching, coalition formation, and gross substitutes. *Econometrica*, 50(6):1483–1504.

Koopmans, T. and Beckmann, M. (1957). Assignment problems and the location of economic activities. *Econometrica*, 25(1):53–76.

Loomes, G. (2005). Modelling the stochastic component of behaviour in experiments: Some issues for the interpretation of data. *Experimental Economics*, 8(4):301–323.

McKelvey, R. and Palfrey, T. (1995). Quantal response equilibria for normal form games. *Games and Economic Behavior*, 10(1):6–38.

Nalbantian, H. and Schotter, A. (1995). Matching and efficiency in the baseball free-agent system: An experimental examination. *Journal of Labor Economics*, 13(1):1–31.

Núñez, M. and Rafels, C. (2003). The assignment game: the τ-value. *International Journal of Game Theory*, 31(3):411–422.

Olson, M. and Porter, D. (1994). An experimental examination into the design of decentralized methods to solve the assignment problem with and without money. *Economic Theory*, 4(1):11–40.

Otto, P. and Bolle, F. (2011). Matching markets with price bargaining. *Experimental Economics*, 14(3):322–348.

Pais, J. and Pintér, Á. (2008). School choice and information: An experimental study on matching mechanisms. *Games and Economic Behavior*, 64(1):303–328.

Pérez-Castrillo, D. and Sotomayor, M. (2002). A simple selling and buying procedure. *Journal of Economic Theory*, 103(2):461–474.

Quint, T. (1991a). Characterization of cores of assignment games. *International Journal of Game Theory*, 19(4):413–420.

Quint, T. (1991b). The core of an m-sided assignment game. *Games and Economic Behavior*, 3(4):487–503.

Quint, T. (1996). On one-sided versus two-sided matching games. *Games and Economic Behavior*, 16(1):124–134.

Roth, A. (1985). Common and conflicting interests in two-sided matching markets. *European Economic Review*, 27(1):75–96.

Rust, J. (2010). Comments on: "structural vs. atheoretic approaches to econometrics" by Michael Keane. *Journal of Econometrics*, 156(1):21–24.

Selten, R. (1972). Equal share analysis of characteristic function experiments. In Sauermann, H., editor, *Contributions to Experimental Economics (Beiträge Zur Experimentellen Wirtschaftsforschung)*, pages 130–165. Mohr Siebeck.

Selten, R. (1991). Properties of a measure of predictive success. *Mathematical Social Sciences*, 21(2):153–167.

Shapley, L. and Shubik, M. (1972). The assignment game i: The core. *International Journal of Game Theory*, 1(1):111–130.

Solymosi, T. and Raghavan, T. (2001). Assignment games with stable core. *International Journal of Game Theory*, 30(2):177–185.

Sotomayor, M. (1999). The lattice structure of the set of stable outcomes of the multiple partners assignment game. *International Journal of Game Theory*, 28(4):567–583.

Stahl, D. and Wilson, P. (1995). On players' models of other players: Theory and experimental evidence. *Games and Economic Behavior*, 10(1):218–254.

Tenbrunsel, A., Wade-Benzoni, K., Moag, J., and Bazerman, M. (1999). The negotiation matching process: Relationships and partner selection. *Organizational Behavior and Human Decision Processes*, 80(3):252–283.

Toda, M. (2005). Axiomatization of the core of assignment games. *Games and Economic Behavior*, 53(2):248–261.

Turocy, T. (2005). A dynamic homotopy interpretation of the logistic quantal response equilibrium correspondence. *Games and Economic Behavior*, 51(2):243–263.

Vuong, Q. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57(2):307–333.

Weizsäcker, G. (2003). Ignoring the rationality of others: evidence from experimental normal-form games. *Games and Economic Behavior*, 44(1):145–171.

Wilcox, N. (2008). Stochastic models for binary discrete choice under risk: A critical primer and econometric comparison. *Risk aversion in experiments*, 12:197–292.

Wilcox, N. (2011). Stochastically more risk averse: A contextual theory of stochastic discrete choice under risk. *Journal of Econometrics*, 162(1):89–104.

Zhang, P. (1993). Model selection via multifold cross validation. *The Annals of Statistics*, 21(1):299–313.