



# Relaxing the symmetry assumption in participation games: a specification test for cluster-heterogeneity

Alan Kirman<sup>1</sup> · François Laisney<sup>2</sup> · Paul Pezanis-Christou<sup>3</sup> 

Received: 20 May 2019 / Revised: 14 February 2023 / Accepted: 19 February 2023 /  
Published online: 5 April 2023  
© The Author(s) 2023

## Abstract

We propose a novel approach to check whether individual behaviour in binary-choice participation games is consistent with the restrictions imposed by symmetric models. This approach allows in particular an assessment of how much cluster-heterogeneity a symmetric model can tolerate to remain consistent with its behavioural restrictions. We assess our approach with data from market-entry experiments which we analyse through the lens of ‘Exploration versus Exploitation’ (EvE, which is equivalent to Logit-QRE) or of Impulse Balance Equilibrium (IBE). We find that when the symmetry assumption is imposed, both models are typically rejected when assuming pooled data and IBE yields more data-consistent estimates than EvE, i.e., IBE’s estimates of session and pooled data are more consistent than those of EvE. When relaxing symmetry, EvE (IBE) is rejected for 17% (42%) of the time. Although both models support cluster-heterogeneity, IBE is much less likely to yield over-parametrised specifications and insignificant estimates so it outperforms EvE in accommodating a model-consistent cluster-heterogeneity. The use of regularisation procedures in the estimations partially addresses EvE’s shortcomings but leaves our overall conclusions unchanged.

**Keywords** Participation games · Exploration versus Exploitation · Logit-Quantal Response Equilibrium · Impulse Balance Equilibrium · Cluster-heterogeneity · Specification test · Regularized minimum distance estimation · Experiments

---

✉ Paul Pezanis-Christou  
paul.pezanis-christou@adelaide.edu.au

Alan Kirman  
kirman@ehess.fr

François Laisney  
laisney@zew.de

<sup>1</sup> École des Hautes Études en Sciences Sociales, CAMS, Paris, France

<sup>2</sup> ZEW – Leibniz-Zentrum für Europäische Wirtschaftsforschung, Mannheim, Germany

<sup>3</sup> School of Economics and Public Policy, University of Adelaide, Adelaide, Australia

**JEL Classification** C7 · C92**1 Introduction**

Over the last three decades, the literature on behavioural economics has proposed a number of models to explain various anomalies that can hardly be organised by the standard equilibrium approach. In the context of games, these models consider alternative preferences, traits and/or rationales which relative explanatory powers have been assessed with laboratory experiments (see e.g., McKelvey & Palfrey, 1995; Selten & Chmura, 2008; Costa-Gomes et al., 2009; Crawford, 2013). While such horseracing approach documents the models' relative goodness-of-fit performances and helps determining a 'best model', it leaves unanswered the question of whether the estimated models are indeed consistent with the restrictions they impose on individuals' behaviour. This article presents a novel approach and a specification test to address this question in the context of symmetric binary-choice participation games such as market-entry games, volunteer's dilemmas, discrete step-level public good and voter participation games. It contributes to the existing literature on this issue in two ways.

First, it provides a useful theory-based selection criterion for models which explanatory powers can hardly be assessed otherwise than by their goodness-of-fit. This is the case with the Quantal Response Equilibrium model (QRE, McKelvey & Palfrey, 1995), a stochastic version of the Nash equilibrium that assumes players to best-respond to their own and to the others' payoff disturbances and which predictions hinge upon the distributional properties of these errors (see Goeree et al., 2016, and the references therein). This model has proven remarkably successful with fitting the data of numerous experiments but its reliance on players' unobservable payoff disturbances has raised concerns about its falsifiability, see Haile et al. (2008). Goeree et al. (2005) addressed such concerns by determining restrictions on these disturbances to bracket QRE's falsifiability (see Goeree et al., 2016, for further discussion on this topic). Golman (2011) deals with this problem in the context of heterogeneous agents and provides conditions under which the behaviour of the representative agent of a pool of individuals may be rationalised by QRE. These conditions determine whether the aggregation of agents' payoff disturbances fulfils the i.i.d. assumption on which QRE builds, and they yield useful predictions for asymmetric binary-choice games by restricting the set of QRE-consistent choice frequencies. On the other hand, Melo et al. (2019) check whether players' behaviour in multiple games is consistent with the QRE hypothesis. Their procedure exploits a set of restrictions on agents' choices in different games and on these games' payoffs. It is also nonparametric in the sense that it does not require the distribution of payoff disturbances in a particular game to be specified.

Unlike these investigations which pertain to QRE settings, ours exploits the behavioural restrictions imposed by a symmetric model on individuals' participation rates only and therefore allows the comparison of different models, including QRE.

Second, it permits an assessment of a model's consistency with the assumption of 'cluster-heterogeneity', whereby individuals with common characteristics (e.g., their participation rates) are clustered together and share a common model-parameter to be estimated. It thus alleviates the problem of modelling heterogeneity, which typically raises questions about which sort of relaxation of "common knowledge" assumption(s) about what agents believe about others can be used and which still allow one to 'close' the model.<sup>1</sup> Rogers et al. (2009), for example, develop QRE models where heterogeneity is modelled either in terms of *common knowledge* beliefs about others' traits (as in Camerer et al., 2016) or of *subjective* beliefs, i.e., each player believes that the others' traits are i.i.d. from the same distribution as her/his own, which is assumed private information (as in Armantier & Treich, 2009).<sup>2</sup>

Although making these modifications shows that assuming heterogeneity considerably improves the model's goodness-of-fit, it also heightens the question of the model's falsifiability since the presumed beliefs about others' behaviour remain difficult to assess. Our approach does not require additional behavioural assumptions about one's own or others' behaviour since it is based on observables, e.g., the players' participation rates; and it allows one to determine how much cluster-heterogeneity a symmetric model can tolerate to remain consistent with the restrictions it imposes on individual behaviour. While the symmetric assumption provides valuable normative predictions for policy recommendations such as the design of markets, contracts and/or bargaining legislations, it is rather unrealistic and thus restrictive. By considering an *observable* cluster-heterogeneity rather than a *hypothetical* heterogeneity in the players' beliefs, we can better assess a model's predictions and possibly broaden its range of applications.

We assess our approach with new data on market-entry games of complete information. These games suit well our case since they involve fairly straightforward incentives and may account for a relatively large number of players (which is needed for studying cluster-heterogeneity). These games have also been widely studied in the social sciences and laboratory experiments typically indicate that participants somehow manage to behave almost optimally since their participation rates often even out the expected profits from entry and from no entry (Ochs, 1990; Sundali et al., 1995; Zwick & Rapoport, 2002).<sup>3</sup> This observation was first coined as 'magic'

<sup>1</sup> This approach is also reminiscent of modifications that have been made to basic modern macroeconomic models to take account of what Angeletos and Lian (2016) describe as "the potential fragility of workhorse macroeconomic models to relaxations of common knowledge". See Kirman (2006) for an overview of the role of heterogeneity in economics, Branch and McGough (2018) for a discussion of the role of heterogeneous expectations in macroeconomics, and Bookstaber and Kirman (2018) on the difficulties with building both theoretical and computational models of heterogeneous agents even when the nature of the heterogeneity is known and understood.

<sup>2</sup> See also McKelvey et al. (2000) and Weizsäcker (2003) for earlier investigations of equilibrium models of games played by non-homogenous agents.

<sup>3</sup> Market-entry games also belong to a broader class of congestion games used to study commuting problems (Selten et al., 2007) and other phenomena such as phase transitions in physics and speculative trading in finance (see Yeung and Zhang, 2009; Challet et al., 2014; Bottazzi and Devetag, 2003). An insightful analysis of a market-entry setting is Arthur's (1994) *El Farol* bar dilemma, which echoes Yogi Berra's famous quote 'Nobody goes there anymore. It's too crowded'.

(Kahneman, 1988; Meyer et al., 1992; Rapoport, 1995), and subsequent experiments have put in perspective the roles of reinforcement learning processes (e.g., Erev & Rapoport, 1998; Rapoport et al., 1998, Duffy & Hopkins, 2005; Erev et al., 2010) and other behavioural traits like probability misperception (Rapoport et al., 2002) and overconfidence (Camerer & Lovo, 1999). Goeree and Holt (2005) examine these and other participation games from a QRE perspective (with a Logit error structure, i.e., the Logit-QRE) and determine conditions to observe under- or over-participation.

We study these market-entry games through the lens of two stationary behavioural models: the ‘Exploration *versus* Exploitation’ dilemma (EvE) outlined in Nadal et al. (1998), Weisbuch et al. (2000), Kirman (2011) and Bouchaud (2013) and which essentially entails a trade-off between maximising current and future profits, or through that of Impulse Balance Equilibrium (IBE, Selten et al., 2005) which balances off the foregone expected payoffs associated to each possible choice. The details of these models are discussed in the next section and we highlight here two of their properties that motivate our experimental investigation. First, EvE is structurally equivalent to Logit-QRE and thus directly relates to the predictions of Goeree and Holt (2005)—in brief, ‘Exploration’ in EvE corresponds to a ‘purely random behaviour’ in QRE, ‘Exploitation’ corresponds to a ‘best-responding behaviour’, and any mix of these two options corresponds to a ‘stochastic best-responding behaviour’. Second, despite their different premises, EvE and IBE fit observed entry probabilities equally well in the range of EvE-consistent choice frequencies. And since the range of IBE-consistent choice frequencies is larger, the usual ‘goodness-of-fit horseshooting’ would document nothing more than occurrences where IBE outperforms EvE and is therefore not pursued.<sup>4</sup> Given these properties, we focus analysis on the models’ relative success with consistently organising behaviour in treatments that manipulate payoff *levels* (i.e., ‘High’ or ‘Low’) and payoff *structures* (i.e., with payoffs from entry depending on attendance in various ways). In addition, we document the sensitivity of our conclusions to the econometric procedures used, i.e., with(out) imposing symmetry, with(out) assuming homoscedastic errors, and with(out) regularisation of the errors’ variance matrix.

We summarise our experimental findings in the following four points. First, imposing symmetry (as is usually done in the literature) yields significant IBE-estimates that are of similar magnitudes across aggregation levels (i.e., session or pooled data) and EvE-estimates that are either insignificant or that bear little consistency across aggregation levels. Second, relaxing symmetry and using OLS estimation methods leads the specification test to reject EvE with cluster heterogeneity less often than IBE no matter the payoff level or structure (17% vs 42% of all sessions). However, when considering the models’ non-rejected specifications, EvE typically yields insignificant

<sup>4</sup> The relative goodness-of-fit performances of these (and other) models has been investigated in the context of 2×2 constant- and non-constant sum games by Selten and Chmura (2008), Brunner et al. (2010), and Selten et al. (2010) who found no clear evidence of a ‘best model’. Brunner et al. also extended the analysis to other binary-choice games to stress-test the particular parametrization of IBE used by Selten and Chmura (2008) but found no evidence of a ‘best model’.

cluster-estimates, and most of its multi-clustered specifications are over-parametrised, i.e., their cluster-estimates are not significantly different from each other. This is not the case for IBE which, in addition, can rationalise the presence or absence of clusters of players with low participation rates. Third, these patterns hardly change when the estimations pertain to the second half of the experiments to account for participants' experience of play. Fourth, when estimating the models with more efficient econometric procedures with(out) regularisation, the EvE-specifications become more likely to be rejected (25% vs 17% of all sessions) and yield less insignificant cluster-estimates. Yet, most of the non-rejected EvE-specifications are still over-parametrised whereas our conclusions for IBE are hardly affected. In sum, our study indicates that IBE yields more consistent estimates than EvE when symmetry is imposed and that it accommodates cluster-heterogeneity better than EvE when it is relaxed.

The next section presents the EvE and IBE models for market-entry games. Section 3 lays out the econometric procedures and our specification test for this class of binary-choice games. The experimental design and procedures are presented in Sect. 4. Section 5 reports the estimation results when symmetry in the players' choices is imposed and when it is relaxed. Section 6 concludes.

## 2 Two stationary models of market-entry games

Assume  $n$  agents who independently decide whether to enter a market or not. Agent  $i$ 's decision is represented by a variable  $d_i$  that takes the value 1 if she enters and 0 if not. The payoff from not entering is constant and equal to  $H$ , whereas the one from entering is a function  $G(\cdot)$  of the number of entrants  $A = \sum_i d_i$ . A congestion problem typically arises if for some integer value  $c < n$ , we have  $G(A) \geq H$  if  $A \leq c$  and  $G(A) < H$  if  $A > c$ . With such a reward scheme, any vector of decisions  $d$  such that exactly  $c$  out of  $n$  agents choose to enter constitutes a pure Nash equilibrium. There are exactly  $\binom{n}{c}$  such equilibria, each yielding an aggregate payoff equal to  $cG(c) + (n - c)H$ .

There may also exist symmetric mixed-equilibrium strategies, i.e., that equalize an agent's expected payoff from entering,  $\pi^E$ , to that from not entering,  $\pi^{NE} = H$ . That is, if  $p$  stands for the common probability of entry, then an equilibrium probability  $p^{Nash}$  solves:

$$\pi^E(p) \equiv \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k} G(k+1) = H \equiv \pi^{NE} \quad (1)$$

where  $k$  is a realization of the random variable  $K$  characterizing the number of entrants other than oneself. Note that (1) requires that the  $n$  agents behave symmetrically in that they all choose to enter with the same probability  $p$ —clearly, one could also consider asymmetric mixed-equilibria in which some agents enter with commonly known probabilities. For reasons that will become clear in Sect. 3, it is

convenient to rewrite this expression as being conditional on  $p_{-i}$ , the  $n - 1$  vector of entry probabilities for agents other than agent  $i$ <sup>5</sup>:

$$\pi_i(d = 1 | p_{-i}) = \sum_{k=0}^{n-1} P[k \text{ go} | p_{-i}] G(k + 1).$$

## 2.1 Exploration versus Exploitation: EvE

In this framework, agents aim at finding a compromise between maximizing their current payoff and keeping themselves informed about market conditions to maximize their future payoffs. In our context, we can think of changing market conditions driven by agents' irregular or stochastic entry behaviour. In this case, agents may find it worthwhile to sometimes explore the alternative option, i.e., entering or not entering the market. While the 'exploitation' part of the dilemma, i.e., the maximization of current payoffs, is straightforward, the 'exploration' part hinges upon the *maximum entropy principle* which captures the agent's information seeking behaviour (see Anderson et al., 1992).<sup>6</sup> In brief, an agent seeking maximal information from her/his decisions would explore each alternative with equal probabilities so that entropy is maximized whereas an agent who does not seek information would clearly avoid exploring and would focus on maximizing current payoffs, so the weight on entropy is minimized. This framework was first used by Nadal et al. (1998) for the study of buyer–seller interactions and we adapt it here for the analysis of market-entry games.

Denote agent  $i$ 's probability of entry by  $p_i$  and that agent's expected payoff from entry in terms of the probabilities of entry of the  $n - 1$  other agents by  $\pi^E(p_{-i})$ . Using Shannon's measure of entropy  $S_i = -p_i \ln p_i - (1 - p_i) \ln (1 - p_i)$  with  $p_i$  neither 0 nor 1, the agent's objective function to maximise is then given by:

$$\begin{aligned} \pi_i &= p_i \pi^E(p_{-i}) + (1 - p_i) H + \sigma S_i \\ &= p_i \pi^E(p_{-i}) + (1 - p_i) H + \sigma [-p_i \ln p_i - (1 - p_i) \ln (1 - p_i)]. \end{aligned} \quad (2)$$

where  $\sigma \geq 0$  is a parameter capturing the weight that agent  $i$  assigns to the preservation of information about market conditions for long term profits. Differentiating this expression with respect to  $p_i$ , we obtain the following first-order condition for maximisation:

$$\pi^E(p_{-i}) - H + \sigma [-\ln p_i + \ln (1 - p_i)] = 0,$$

or equivalently (with  $\sigma = 1/\lambda$ )

<sup>5</sup> See Appendix 1 for the expression of the probability  $P[k \text{ go} | p_{-i}]$  that  $k$  agents enter given the vector  $p_{-i}$ .

<sup>6</sup> For a discussion of this principle in economics and game theory see Scharfenaker and Yang (2020).

$$p_i = \frac{1}{1 + \exp\{-\lambda[\pi^E(p_{-i}) - H]\}}. \quad (3)$$

This yields a system of  $n$  equations if there are  $n$  agents, and given the homogenous weighting parameter  $\lambda$ , this should be solved for the vector  $p^* = (p_1, p_2, \dots, p_n)$ . Under the assumption of symmetry,  $p_{-i}$  has all its components equal to  $p_i$ , which we simply denote by  $p$ , and thus  $p$  and  $\lambda$  are related by: or equivalently

$$p = \frac{1}{1 + \exp\{-\lambda[\pi^E(p) - H]\}}$$

$$\lambda = \frac{\ln \frac{p}{1-p}}{\pi^E(p) - H}. \quad (4)$$

Note that this exactly matches McKelvey and Palfrey's definition of a Logit-QRE (with  $\lambda$  standing for the agents' homogenous 'best-responsiveness') so that the models are structurally equivalent if agents' payoff shocks in QRE are extreme-value i.i.d and if EvE assumes Shannon's entropy measure.<sup>7</sup> Thus, if rational agents behave symmetrically and do not explore, then  $p$  is such that  $\pi^E(p) = H$ , i.e.,  $p = p^{Nash}$  and  $\lambda \rightarrow \infty$ . On the other hand, if they maximise exploration, then they choose  $p$  such that  $p = 1 - p = 0.5$ , so that  $\lambda \rightarrow 0$ . If  $p > 0.5$ ,  $\lambda$  is positive if  $\pi^E(p) > H$  and it is negative (theory-inconsistent) otherwise. The Maximum Likelihood estimate of  $p$ , assuming independent observations, is the relative frequency of entry,  $\bar{d}_n$ , and the Maximum Likelihood Estimator (MLE) of  $\lambda$  follows from (4). Note that  $\bar{d}_n$  remains a statistically consistent estimator for  $E(d) = p$  for less restrictive covariance structures of the observations, by various flavours of the weak laws of large numbers.

## 2.2 Impulse Balance Equilibrium: IBE

IBE basically assumes that if at some stage an alternative option would have yielded a higher payoff, then the agent receives an impulse to use this alternative in the next stage, i.e., agents only take account of foregone payoffs, as in Learning Direction Theory (Selten & Buchta, 1999). It is defined as the long run outcome of such stage-to-stage behaviour. In the context of market-entry games, an agent receives an impulse for entry if the payoff received from not entering is smaller than that from entering. Denoting by  $I$  the number of other entrants and

<sup>7</sup> See Appendix 2 for the derivation of the Logit-QRE for this game. The i.i.d. assumption is central to QRE and reverts to assuming that agents take their decisions independently and do not interact with each other in EvE models. The modelling of dynamics in settings with time-correlated decisions and heterogeneous agents quickly becomes intractable and the determination of equilibria is confined to special cases, see Bouchaud (2013) and Goeree et al. (2016). EvE has also been formalised in terms of 'rational inattention' by Matějka and McKay (2015), see Gabaix (2018) for a review of this literature. See also Evans and Prokopenko (2021) for an application of EvE with a state dependant variant of Shannon's measure to study housing market data.

by  $p$  the common probability of entering the market, the expected magnitude of these impulses for entry is defined as:

$$\begin{aligned} IMP^E(p) &= E[G(K+1)\mathbb{I}_{\{G(K+1)>H\}}] \\ &= \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k} G(k+1) \mathbb{I}_{\{G(k+1)>H\}} \end{aligned}$$

or equivalently in terms of  $p_{-i}$  rather than  $p$ :

$$IMP^E(p_{-i}) = \sum_{k=0}^{n-1} P[k \text{ go} | p_{-i}] G(k+1) \mathbb{I}_{\{G(k+1)>H\}}. \quad (5)$$

Similarly, an agent receives an impulse for no entry if the payoff received from entering is not larger than that from not entering. The expected magnitude of these impulses for no entry is defined as:

$$\begin{aligned} IMP^{NE}(p) &= H \cdot P[G(K+1) < H] \\ &= H \left[ 1 - \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k} \mathbb{I}_{\{G(k+1)>H\}} \right] \end{aligned}$$

or equivalently

$$IMP^{NE}(p_{-i}) = H \left[ 1 - \sum_{k=0}^{n-1} P[k \text{ go} | p_{-i}] \mathbb{I}_{\{G(k+1)>H\}} \right]. \quad (6)$$

Note that these impulses are defined relatively to the game's maximin pure strategy of not entering the market which yields a sure payoff of  $H$ . Selten and Chmura (2008) further observe that receiving a payoff lower than this sure payoff should be perceived as a loss. To this extent, and in the light of empirical and experimental evidence of loss aversion in agents' preferences (Bernatzi & Thaler, 1995; Tversky & Kahneman, 1991), we follow Ockenfels and Selten (2005) and define an IBE for this market-entry game such that agent  $i$  is indifferent between 'receiving  $IMP^E(p_{-i})$  and entering' and 'receiving  $\kappa IMP^{NE}(p_{-i})$  and not entering', where  $\kappa > 0$  stands for an impulse weight. That is, agent  $i$  would choose to enter the market with probability  $p_i$  that equalises her expected weighted impulses:

$$p_i IMP^E(p_{-i}) = (1 - p_i) \kappa IMP^{NE}(p_{-i}) \quad (7)$$

This impulse balance equation characterizes a long-run IBE in which participants do no more react to the expected impulses they receive. We could of course consider a short-run IBE, i.e., that would solve  $IMP^E(p_{-i}) = \kappa IMP^{NE}(p_{-i})$ , but the resulting IBE for agent  $i$  would then be independent of  $p_i$  and, as shown in the next section, this would considerably limit the scope of our study.



Finally, unlike Selten and Chmura (2008) who assume  $\kappa = 2$ , we estimate the impulse weight  $\kappa$  by Maximum Likelihood (as for EvE) so the estimator of  $p$  is  $\bar{d}_n$  and the MLE of  $\kappa$  (assuming symmetry and  $p \neq 1$ ) follows from

$$\kappa = \frac{pIMP^E(p)}{(1-p)IMP^{NE}(p)}. \quad (8)$$

### 3 A specification test: the $\Sigma$ -test

When we assume symmetry, the models we consider only propose a reparametrization  $\lambda(p)$  for EvE and  $\kappa(p)$  for IBE. Thus, under symmetry, there is no scope for discriminating between these models beyond commenting on implausible values of  $\lambda(p)$  and  $\kappa(p)$ . If we do not impose symmetry, then (3) and (7) can be rewritten as systems of linear restrictions on parameters  $\lambda$  and  $\kappa$ :

$$\ln \frac{p_i}{1-p_i} - \lambda[\pi_i^E(p_{-i}) - H] = 0 \quad \text{for } i = 1, \dots, n \quad (9)$$

and

$$p_i IMP^E(p_{-i}) - (1-p_i)\kappa IMP^{NE}(p_{-i}) = 0 \quad \text{for } i = 1, \dots, n \quad (10)$$

Both systems can thus be written in the form  $y(p) - \theta x(p) = g(p, \theta) = 0$ , with  $\theta = \lambda$  or  $\kappa$ , and with  $y$ ,  $x$  and  $g$  vector functions with values in  $\mathbb{R}^n$ . The proposed formulation of (9) and (10) in terms of  $p_{-i}$  makes it possible to express the EvE or IBE model for homogenous players—in the sense that they share a common single parameter—while still allowing for possibly different individual entry probabilities, and to design a specification test. A further possibility we shall explore is to allow for cluster-heterogeneous players, i.e., players with similar characteristics (e.g., entry-probabilities) whom the model considers identical by assigning them the same parameter. In this case,  $\theta$  is a vector instead of a scalar and the length of the vector directly affects the power of the  $\Sigma$ -test since a vector of length  $n$  represents full heterogeneity and leads to never rejecting the null of consistency.

Given the asymptotically normal estimator  $\hat{p}_T$  of  $p$ , the vector of individual entry frequencies, with asymptotic variance  $V$  of which we describe a consistent estimator  $\hat{V}_T$  in Appendix III.A, an optimal asymptotic least squares estimator of  $\theta$  is<sup>8</sup>:

$$\begin{aligned} \hat{\theta}_T &= \arg \min_{\theta} g'(\hat{p}_T, \theta) \hat{S}_T^{-1} g(\hat{p}_T, \theta) \\ &= [x'(p) \hat{S}_T^{-1} x(p)]^{-1} x'(p) \hat{S}_T^{-1} y(p) \end{aligned} \quad (11)$$

with  $\hat{S}_T$  converging to

<sup>8</sup> See e.g. Gouriéroux and Monfort (1995), for a justification of all the assertions in this section.

$$S = \frac{\partial g(p, \theta)}{\partial p'} V \frac{\partial g'(p, \theta)}{\partial p}.$$

$\hat{\theta}_T$  is thus the GLS estimator in the regression of  $y(p)$  on  $x(p)$ , the variance of the error term being  $S$ .

Given a preliminary estimate of  $\theta$ , say  $\tilde{\theta}_T$  obtained by replacing  $\hat{S}_T$  in (11) with the identity matrix, i.e.,  $\tilde{\theta}_T$  is the OLS estimator in the regression of  $y(p)$  on  $x(p)$ , a consistent estimator of  $S$  is:

$$\hat{S}_T(\hat{p}_T, \tilde{\theta}_T) = \frac{\partial g(\hat{p}_T, \tilde{\theta}_T)}{\partial p'} \hat{V}_T \frac{\partial g'(\hat{p}_T, \tilde{\theta}_T)}{\partial p}.$$

The asymptotic variance of  $\hat{\theta}_T$  is given by  $V_{asy}(\hat{\theta}_T) = [x'(p)S^{-1}x(p)]^{-1}$  and a consistent estimator is  $\widehat{V_{asy}}(\hat{\theta}_T) = [x'(\hat{p}_T)\hat{S}_T^{-1}x(\hat{p}_T)]^{-1}$ . Under the null that there exists  $\theta$  such that  $g(p, \theta) = 0$  for the true  $p$ , or in other words that the restrictions on entry probabilities embodied by the model are valid,

$$Tg'(\hat{p}_T, \hat{\theta}_T)\hat{S}_T^{-1}[\hat{S}_T(\hat{p}_T, \tilde{\theta}_T)]^{-1}g(\hat{p}_T, \hat{\theta}_T) \approx \chi^2(n-1), \quad (12)$$

and this over-identification test can be used to test the underlying theory. All we need for the implementation of this specification test, for short the  $\Sigma$ -test, are thus  $\hat{V}_T$  and the derivatives  $\partial g_i(p, \theta)/\partial p_i$ . The technical details for the determination of these expressions are given in Appendix III.B. The number of degrees of freedom is  $n-1$  when assuming homogeneity [i.e., the length of vector  $\theta$  is 1, cf. (12)] and it is at most  $n-K$  when assuming heterogeneous players sorted in  $K$  clusters (i.e., the length of vector  $\theta$  is  $K$ ), as discussed in Appendix III.C.

Note finally that since this test exploits the game's probabilistic structure by rewriting agents' probabilities of entry as a function of  $p_{-i}$ , it can be tailored for the assessment of behaviour in other binary-choice participation games like the volunteer's game, the (discrete) step-level public good game and voter participation games. This, of course, remains conditional on having well-defined predictions to test, as is the case for EvE and QRE in general but not necessarily for IBE since its long-run equilibrium may not always be defined.<sup>9</sup>

## 4 Experimental design and procedures

The experiments involve groups of 10 participants and a  $2 \times 3$  factorial design which assumes two payoff *levels*, High and Low, and three payoff *structures*: one two-step payoff function (DISC) yielding a positive payoff  $G$  from entering if attendance

<sup>9</sup> In the Volunteer's dilemma game, for example, players receive a gain  $G$  if at least one of them incurs the cost  $C < G$  of volunteering. Assuming symmetric agents, the expected impulses from 'volunteering' and 'not volunteering' are then defined as  $IMP^V(p) = (1-p)^{n-1}(G-C)$  and  $IMP^{NV}(p) = [1 - (1-p)^{n-1}]G$ , respectively, so the long-run IBE would solve the equation  $pIMP^V(p) = (1-p)\kappa IMP^{NV}(p)$  which solution  $p \in (0, 1]$  may not exist for some values of  $\kappa \geq 0$ .

$A < c$  and 0 otherwise, and two non-monotone ones (NOM1 and NOM2) in which payoffs first increase and then decrease with  $A$ . The binary payoff structure of DISC implies that the players' choices are strict substitutes whereas the non-monotone structures introduce both strategic complementarity and strategic substitutability in the players' actions that have been theoretically studied in the context of global congestion games (see e.g., Karp et al., 2007) but which effects in complete information settings have not yet been investigated experimentally.<sup>10</sup>

These payoff structures are displayed in Fig. 1, and the models' equilibrium relationships between  $p$  and  $\lambda > 0$  or  $\kappa > 0$  for the treatments considered are shown in Fig. 2. For each payoff level, both DISC and NOM1 yield  $\binom{10}{6} = 210$  Nash equilibria in pure strategies, unique mixed-equilibrium strategies and unique IBE strategies whereas NOM2 has one more equilibrium in pure strategies (where all agents choose not to enter), two mixed-equilibrium strategies and two IBE strategies (one with a low entry-probability and one with a high entry-probability).<sup>11</sup>

We are interested in checking if and how behaviour is affected by these payoff structures and to what extent it is consistent with EvE and/or IBE when allowing for cluster-heterogeneity. In this regard, since the ranges of probabilities for which EvE and IBE yield model-consistent estimates in DISC and NOM1 are  $[0.5, p^{Nash}]$  for EvE and  $[0, 1]$  for IBE, the models' cluster-estimates should lie within these ranges and be significantly different from each other for cluster-heterogeneity to be model-consistent and significant. It thus follows that the scope for IBE to accommodate the latter in these treatments is considerably larger than that for EvE.<sup>12</sup>

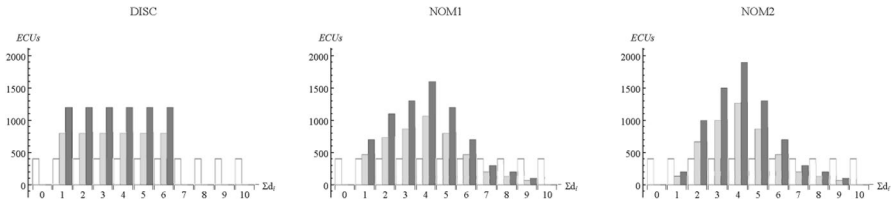
A similar argument holds for NOM2 since the mixed-equilibria have different loci of consistent choice frequencies (defined either on  $[0.5, p_1^{Nash}]$  or on  $[0, p_2^{Nash}]$  with  $p_2^{Nash} < 0.5$ ) whereas the IBE equilibria have a unique locus (because both equilibria depend on a common  $\kappa$ ), so the identification of model-consistent clusters of participants playing in such different (Nash or IBE) equilibria can be achieved with IBE but not with EvE.

Our motivation to consider different payoff levels is to check whether the payoffs' magnitude affects the presence of model-consistent clusters of players, and thus to possibly complement the findings of McKelvey et al. (2000) who report no significant payoff-magnitude effect on the participants' QRE best-responsiveness in  $2 \times 2$  games and evidence of a heterogeneous play.

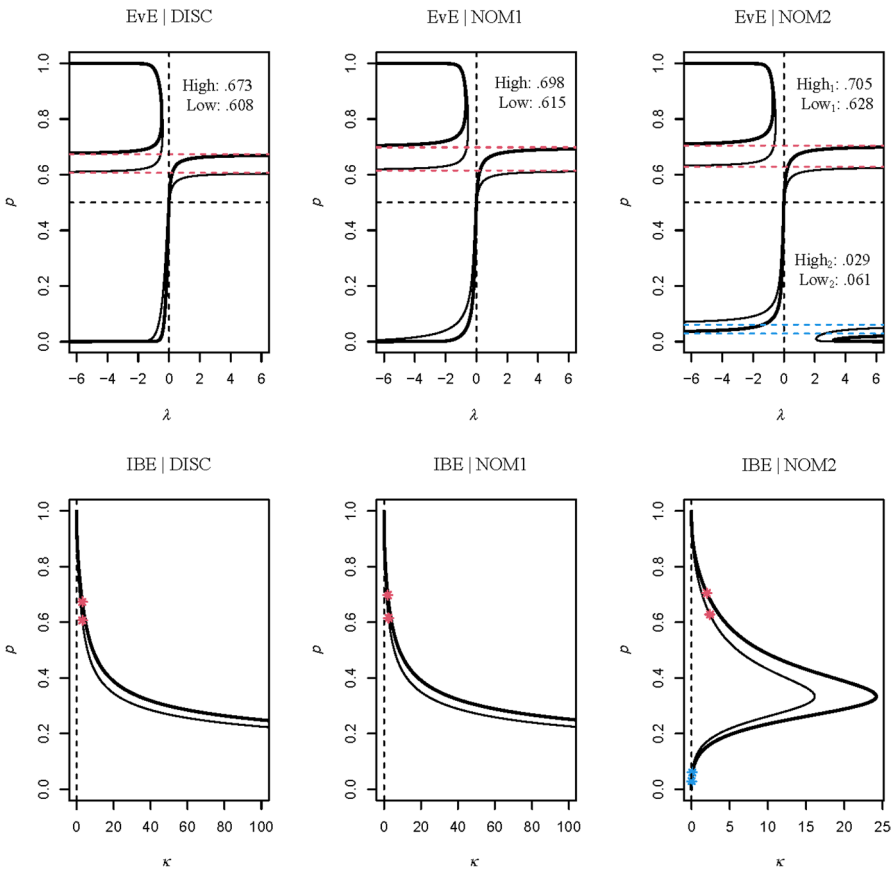
<sup>10</sup> See also Anderson and Engers (2007) who study the equilibria of a class of participation games which payoffs from entry are either monotone increasing or monotone decreasing.

<sup>11</sup> Note that the loci of EvE-consistent choice frequencies in Fig. 2 are such that  $p^{Nash}$  is (almost) reached at values of  $\lambda$  well below infinity. This results from EvE's sensitivity to the magnitude of agents' expected payoff differences  $\pi^E(p_{-i}) - H$ . The plots (and our estimations) assume the payoff figures of Appendix IV.A to be divided by 100 and we observe that scaling them further down would 'flatten' the loci without changing  $p^{Nash}$ , i.e.,  $p^{Nash}$  would be (almost) reached at higher values of  $\lambda$ , but EvE's rationale would then refer to smaller expected payoff differences.

<sup>12</sup> It also follows that when symmetry is assumed, the models yield the same goodness-of-fit for the range of entry-probabilities that are consistent with EvE and that IBE yields a better fit in all treatments otherwise.



**Fig. 1** Payoff levels and structures of market-entry games. No filling stands for ‘No Entry’, light gray (dark gray) stands for ‘Entry’ when payoffs are Low (High). Payoffs expressed in Experimental Currency Units—see Appendix IV.D for exact figures



**Fig. 2** Relationship between  $p$  and EvE's  $\lambda$  or IBE's  $\kappa$ . Thick (Thin) lines stand for High (Low) payoff levels. For EvE, the plots report the  $p^{Nash}$  predictions for each payoff structure and level (cf. coloured horizontal lines). For IBE, the plots display the  $p^{Nash}$  predictions (cf. dots) for each payoff structure and level. As  $\kappa \rightarrow \infty$ ,  $p \rightarrow 0$  in DISC and NOM1

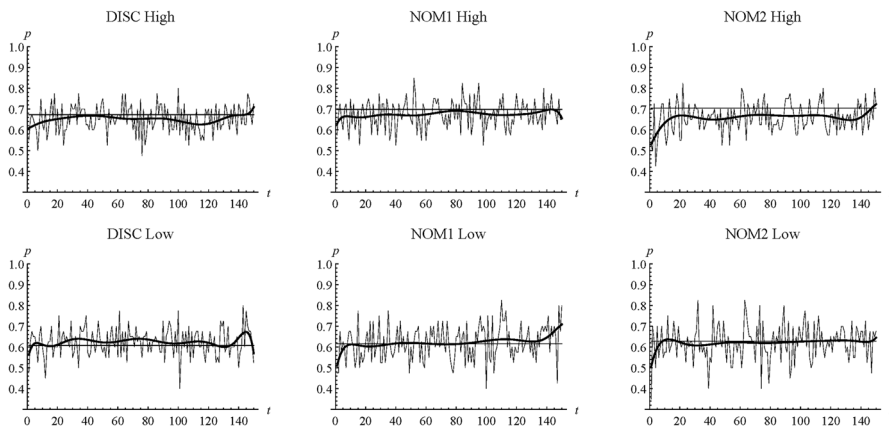
The experiments were conducted at the Laboratory for Experimental Economics of the University of Jaume I (Spain). Participants were undergraduate students in Business Administration, Law or Engineering and were recruited by public advertisement on campus. We conducted eight sessions per payoff structure (DISC, NOM1, NOM2) with 10 participants per session, totalling 240 individuals. For each payoff structure, we conducted four sessions with Low payoffs and four sessions with High payoffs. The experiments were conducted with a between-subject matching protocol and participants could play in only one session. Upon arriving in the laboratory, they were randomly assigned to cubicles equipped with computer terminals and were given instructions that were read aloud.<sup>13</sup> To avoid framing effects, we presented the game in neutral language by asking participants to choose between actions A and B. Each session involved 150 rounds of play, and at the end of each round, participants were only informed about the total number of players in their group who chose B ("No entry"), their own payoff in that round and their cumulated payoff. This information was appended to a "History" window that could be seen at any time during the experiment. Although participants played in fixed groups of 10, we believe that the provision of a sparse end-of-round information feedback combined with the relatively large number of players (10), and a relatively large 'market size-to-capacity' ratio (60%) renders entry-coordination very difficult to achieve. Each session lasted a maximum of 1 h, including the time needed to read the instructions. Participants were rewarded for each round of play at the rate of 0.02 € per 100 points and individual average earnings were €12.77 (i.e., €11.94 in the Low payoff sessions and €13.60 in High payoff ones).

## 5 Results

We start with an overview of the data by displaying the evolution of averaged entry probabilities and their polynomial fits in Fig. 3. The plots suggest an under-entry ( $\hat{p} < p^{Nash}$ ) in all High payoff treatments, and that the 'magic' ( $\hat{p} \approx p^{Nash}$ ) is more likely to hold when payoffs are Low, especially in NOM1 and NOM2. These entry patterns are also present in the session data (cf. Appendix V) and in line with the session and treatment average entry rates of Table 1.

The treatment (pooled) figures of Table 1 show no support for the predicted ranking of entry rates  $p_{DISC}^{Nash} < p_{NOM1}^{Nash} < p_{1,NOM2}^{Nash}$ . Pairwise comparisons indicate a substantially higher entry rate in NOM1 than in DISC and NOM2 when payoffs are High and similar entry rates when they are Low. They also significantly increase with the payoff level, as predicted in equilibrium and as reported by Zwick and Rapoport (2002) who study the effect of 'low' and 'high' entry costs in treatments with a similar 'market size-to-capacity' ratio (50%). We summarise this overview of the pooled data as follows:

<sup>13</sup> The experiments were conducted with the *z-Tree* software (Fischbacher, 2007). See Appendix IV for an English transcript of the set of instructions.



**Fig. 3** Evolution of average probabilities of entry. Horizontal lines stand for the symmetric mixed-equilibrium predictions (we only consider the high-probability equilibrium of NOM2). Bold lines represent polynomial fits of degree 10

**Table 1** Average entry probabilities

| Level | Structure<br>$p^{Nash}$ | Session 1                         | Session 2                          | Session 3                          | Session 4                          | Pooled                             |
|-------|-------------------------|-----------------------------------|------------------------------------|------------------------------------|------------------------------------|------------------------------------|
| High  | DISC                    | <b>.656</b><br><i>.673</i>        | .647 <sup>b</sup><br>[.622, .671]  | <b>.651</b><br><i>[.617, .685]</i> | <b>.655</b><br><i>[.632, .678]</i> | .652 <sup>b</sup><br>[.645, .659]  |
|       | NOM1                    | .669 <sup>b</sup><br>[.640, .697] | .658 <sup>b</sup><br>[.619, .669]  | <b>.677</b><br><i>[.652, .702]</i> | <b>.677</b><br><i>[.655, .699]</i> | .676 <sup>b</sup><br>[.669, .681]  |
|       | NOM2                    | .669 <sup>b</sup><br>[.644, .695] | .635 <sup>b</sup><br>[.609, .660]  | .660 <sup>b</sup><br>[.639, .681]  | .673 <sup>b</sup><br>[.653, .693]  | .659 <sup>b</sup><br>[.653, .665]  |
|       |                         |                                   |                                    |                                    |                                    |                                    |
|       |                         |                                   |                                    |                                    |                                    |                                    |
| Low   | DISC                    | <b>.630</b><br><i>.607</i>        | <b>.595</b><br><i>[.577, .614]</i> | .650 <sup>a</sup><br>[.622, .678]  | <b>.621</b><br><i>[.597, .646]</i> | .624 <sup>a</sup><br>[.618, .630]  |
|       | NOM1                    | <b>.629</b><br><i>.615</i>        | <b>.643</b><br><i>[.612, .674]</i> | <b>.620</b><br><i>[.595, .645]</i> | <b>.594</b><br><i>[.566, .622]</i> | <b>.622</b><br><i>[.615, .628]</i> |
|       | NOM2                    | <b>.628</b><br><i>.628</i>        | .602 <sup>b</sup><br>[.581, .623]  | <b>.619</b><br><i>[.596, .643]</i> | <b>.634</b><br><i>[.604, .664]</i> | .621 <sup>b</sup><br>[.615, .627]  |
|       |                         |                                   |                                    |                                    |                                    |                                    |
|       |                         |                                   |                                    |                                    |                                    |                                    |

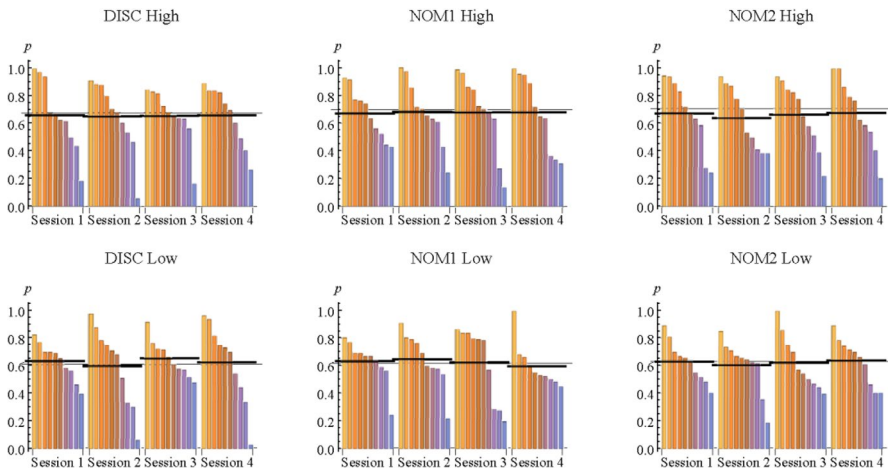
Each 'session' ('pooled') estimate refers to 1500 (6000) observations; Nash mixed-equilibrium predictions in italics; bold cells characterize instances where the symmetric mixed-equilibrium strategy cannot be rejected at the 5% level; 95% Confidence Intervals (based on Newey–West variance estimates) in brackets

<sup>a</sup>Significant over-entry, i.e., when  $p^{Nash}$  is smaller than the lower bound of the 95% CI

<sup>b</sup>Significant under-entry, i.e., when  $p^{Nash}$  is greater than the upper bound of the 95% CI

Observation 0: (A) There is under-entry when payoffs are High. When payoffs are Low, there is (1) over-entry in DISC, (2) a weak support for the Nash mixed-equilibrium play in NOM1, and (3) under-entry (with respect to the high probability equilibrium) in NOM2.

(B) The effect of the payoff structure is most salient when payoffs are High and yields a substantially higher average entry rate in NOM1. Average entry



**Fig. 4** Bar-charts of individual probabilities of entry. Each vertical bar represents an individual. Horizontal thin (thick) lines stand for the symmetric mixed-equilibrium predictions (average probabilities of entry)

rates also significantly increase with the payoff level, as expected in equilibrium.

Before estimating the models, we briefly assess the symmetry of individuals' entry probabilities. The bar-charts in Fig. 4 reveal minor differences in average entry probabilities *between* the sessions of a treatment, and large *within*-session disparities with clusters of participants displaying a similar entry behaviour.<sup>14</sup> The data also show no support for the 'low probability' mixed-equilibrium of NOM2 so we will always refer to the 'high probability' equilibrium of this treatment when discussing our estimation results.

### 5.1 Structural estimations when imposing symmetry

Table 2 reports the (pseudo-)Maximum Likelihood estimation outcomes of EvE and IBE when assuming symmetric players and unknown forms of autocorrelation and heteroskedascity in the errors. As the log-likelihood values contain no information about the model's goodness-of-fit beyond the estimated probability of entry  $\hat{p}$ , we focus on the estimates' overall consistency with Observation 0, and on their

<sup>14</sup> This clustering of players is confirmed in the last 75 rounds (see Appendix VI) and relates to the prediction of Duffy and Hopkins (2005) that participants eventually learn (by reinforcement) to use the pure strategy of always or never entering. The authors find support for this prediction in market-entry experiments with a capacity of two only if the (six) competitors' choices and identities are disclosed after each round of play, which is not the case in our experiments. We report on the participants entry-behaviour in the last 75 rounds and on simple learning trends in the next section.

**Table 2** EvE and IBE estimates

| EVE ESTIMATES ( $\lambda$ ). |           |                                                    |                         |                         |                         |                          |
|------------------------------|-----------|----------------------------------------------------|-------------------------|-------------------------|-------------------------|--------------------------|
| Level                        | Structure | Session 1                                          | Session 2               | Session 3               | Session 4               | Pooled                   |
| High                         | DISC      | 1.258<br>[-.77, 3.29]                              | .762<br>[-.08, 1.60]    | .924<br>[-.70, 2.55]    | 1.205<br>[-.54, 2.95]   | 1.000<br>[.63, 1.37]     |
|                              | NOM1      | 1.044<br>[-.71, 2.27]                              | 2.023<br>[-1.82, 5.87]  | 1.521<br>[-.51, 3.56]   | 1.577<br>[-.36, 3.51]   | 1.430<br>[.98, 1.88]     |
|                              | NOM2      | .812<br>[.07, 1.56]                                | .305<br>[.12, .49]      | .595<br>[.22, .97]      | .921<br>[.19, 1.65]     | .580<br>[.48, .68]       |
| Low                          | DISC      | -1.139<br>[-2.28, .00]                             | 1.660<br>[-1.33, 4.65]  | -.711<br>[-1.03, -.39]  | -1.698<br>[-4.21, .82]  | -1.452<br>[-1.91, -1.00] |
|                              | NOM1      | -2.191<br>[-5.58, 1.20]                            | -1.236<br>[-2.29, -.18] | -6.006<br>[-36.7, 24.7] | 1.016<br>[-.67, 2.70]   | -4.504<br>[-9.05, .041]  |
|                              | NOM2      | 302.621<br>[<-10 <sup>4</sup> , 10 <sup>4</sup> >] | .806<br>[-.04, 1.66]    | 2.866<br>[-5.42, 11.2]  | -4.864<br>[-28.6, 18.9] | 3.509<br>[.28, 6.74]     |
| IBE ESTIMATES ( $\kappa$ ).  |           |                                                    |                         |                         |                         |                          |
| Level                        | Structure | Session 1                                          | Session 2               | Session 3               | Session 4               | Pooled                   |
| High                         | DISC      | 3.449<br>[2.92, 3.98]                              | 3.652<br>[3.11, 4.20]   | 3.564<br>[2.82, 4.31]   | 3.463<br>[2.97, 3.95]   | 3.531<br>[3.39, 3.68]    |
|                              | NOM1      | 2.511<br>[1.99, 3.04]                              | 1.836<br>[1.53, 2.15]   | 2.366<br>[1.92, 2.81]   | 2.354<br>[1.96, 2.75]   | 2.387<br>[2.28, 2.50]    |
|                              | NOM2      | 2.662<br>[2.15, 3.17]                              | 3.445<br>[2.80, 4.09]   | 2.856<br>[2.42, 3.30]   | 2.595<br>[2.20, 3.00]   | 2.874<br>[2.75, 3.00]    |
| Low                          | DISC      | 2.693<br>[2.24, 3.15]                              | 3.306<br>[2.94, 3.67]   | 2.386<br>[1.97, 2.80]   | 2.836<br>[2.43, 3.25]   | 2.790<br>[2.68, 2.89]    |
|                              | NOM1      | 2.224<br>[1.84, 2.61]                              | 2.013<br>[1.57, 2.46]   | 2.374<br>[1.96, 2.79]   | 2.846<br>[2.29, 3.41]   | 2.347<br>[2.23, 2.46]    |
|                              | NOM2      | 2.412<br>[1.99, 2.84]                              | 2.908<br>[2.46, 3.35]   | 2.568<br>[2.14, 3.00]   | 2.309<br>[1.80, 2.82]   | 2.541<br>[2.43, 2.66]    |

Each 'session' ('pooled') estimate refers to 1500 (6000) observations; 95% CIs based on Newey–West variance estimates in brackets; shaded cells characterize instances where the symmetric mixed-equilibrium strategy cannot be rejected at  $\alpha = 5\%$ , cf. Table 1; (dashed-)framed cells characterize (almost) significantly negative estimates; italicised figures indicate insignificant estimates  $\alpha = 5\%$ , i.e., maximal exploration in EvE.

*data-consistency*, i.e., that a treatment's session estimates are of similar magnitude and significance as the estimate for the pooled data.<sup>15</sup>

Looking first at the outcomes for EvE, it appears that except for NOM2/High, all sessions report insignificant or inconsistent (negative) estimates no matter if  $p^{Nash}$  is rejected or not (cf. shaded cells) or if their average entry rates indicate under-entry (cf. Table 1 and Fig. 4). Such insignificant estimates support maximal exploration whereas inconsistent ones result from EvE's inability to rationalize over-entry when  $p^{Nash} > 0.5$ ,

<sup>15</sup> The log-likelihood values are uninformative because they are defined as  $T[\hat{p}\ln\hat{p} + (1 - \hat{p})\ln(1 - \hat{p})]$ , with  $T$  standing for the number of observations. Further, since our estimations account for unknown forms of auto-correlation and heteroscedasticity in the observations, the models' log-likelihoods would be of doubtful interest.



as shown in Fig. 2. In the case of NOM2/High, they are all significantly positive and support a contained exploitation that is in line with the observed under-entry.

The pooled EvE-estimates indicate a contained exploitation in all High payoff structures and in NOM2/Low, and they are otherwise inconsistent (or almost so) as a result of over-entry. Thus, besides a significant under-entry in NOM2/High, the EvE-estimates provide no evidence of a data-consistent behaviour when the estimations impose a symmetric play.

This sharply contrasts with the outcomes for IBE since the session estimates are all significantly positive, typically larger when payoffs are High in DISC and NOM2, and similar across payoff levels in NOM1. This is confirmed by the treatments' estimates which pairwise-comparisons further indicate that  $\hat{\kappa}_{DISC} > \hat{\kappa}_{NOM2} > \hat{\kappa}_{NOM1}$  when payoffs are High and  $\hat{\kappa}_{DISC} > \hat{\kappa}_{NOM1} \approx \hat{\kappa}_{NOM2}$  otherwise. We summarise the above in the following observation:

Observation 1: When assuming symmetric players and estimating the models with pseudo-Maximum Likelihood methods:

(A) The EvE-estimates are data-consistent in NOM2/High and indicate a contained exploitation that is in keeping with the observed under-entry. Otherwise, they are data-inconsistent: they mostly indicate maximal exploration whereas pooled estimates are either negative (thus inconsistent) or they support a contained exploitation.

(B) The IBE-estimates are data-consistent and in keeping with Observation 0. They indicate:

- (1)  $\hat{\kappa}_{High} > \hat{\kappa}_{Low}$  in DISC and NOM2, and  $\hat{\kappa}_{High} \approx \hat{\kappa}_{Low}$  in NOM1.
- (2)  $\hat{\kappa}_{DISC} > \hat{\kappa}_{NOM2} > \hat{\kappa}_{NOM1}$  when payoffs are High and  $\hat{\kappa}_{DISC} > \hat{\kappa}_{NOM2} \approx \hat{\kappa}_{NOM1}$  when they are Low.

## 5.2 Structural estimations when relaxing symmetry

We now estimate the models without imposing symmetry and we run our specification test to assess the consistency of estimates with the restrictions that either model imposes on individual behaviour. Note that the  $\Sigma$ -test only suits the analysis of session data, i.e., games with  $n$  players.

For each session, we cluster the entry probabilities  $p_i$  using the *kmeans* procedure (with 20 random initial values) and estimate each model and its inverse form with  $K = \{1, 2, 3, 4\}$  clusters; each cluster having its own  $\theta$ -parameter (where  $\theta$  is either to  $\lambda$  or  $\tau$ ).<sup>16</sup> This generates eight specifications for each model and

<sup>16</sup> We use the inverse regressions  $x(p)$  on  $y(p)$  (cf. Sect. 3) as a convergence check, since both the direct and inverse approaches must yield the same optimum when we optimize the test statistic (12). Also, we do not consider specifications with more than four clusters because the high volatility and imprecision of parameter estimates when there are four clusters do not encourage to go further. The technical details to determine the agents' clusters based on their entry probabilities are provided in Appendix III.C—the alternative to cluster the  $(x, y)$  vectors defined below Eq. (10) would have led to different groupings for IBE and QRE and was therefore not pursued.

treatment which we estimate with OLS procedures. For each session, model (IBE and EVE) and value of  $K$ , we select the ‘best’ specification in terms of the estimates’ theoretical consistency and the credibility of their confidence intervals. Next, for each session and model, we select the estimated specification with the smallest number of clusters,  $K_{Min}$ , needed to not reject the  $\Sigma$ -test at  $\alpha = 5\%$ . Thus, the reported estimation results document the models’ non-rejections of the  $\Sigma$ -test when  $K_{Min} < 4$ , and their rejections or non-rejections when  $K_{Min} = 4$ . Noting that a rejection with  $K_{Min} = 4$  can reasonably be seen as disqualifying the model when  $n = 10$ , we focus discussion on specifications that do not reject the  $\Sigma$ -test.

The estimation outcomes are relegated to Tables VII.A.1–4 in Appendix VII.A, and since they display no obvious pattern in terms of payoff structure, we start with summarising their main characteristics for each payoff level in the upper panel of Table 3. The first three columns tally the models’ rejections and non-rejections of the  $\Sigma$ -test when  $K_{Min} = 1$  (i.e., homogeneity is not rejected) or when  $1 < K_{Min} \leq 4$  (i.e., homogeneity is rejected in favour of cluster-heterogeneity).

EvE is not rejected for a total of 20 sessions (out of 24, 83%) whereas IBE is not rejected for a total of 14 sessions (58%). Of these non-rejected specifications, EvE supports cluster-heterogeneity in 12 sessions (60%) whereas all non-rejected

**Table 3** Summary of specification test outcomes: OLS procedures

|                       | #(Rejections) | #(Non-rejections) |                      |  | Details  |                       |                                             |
|-----------------------|---------------|-------------------|----------------------|--|----------|-----------------------|---------------------------------------------|
|                       |               | $K_{Min} = 1$     | $1 < K_{Min} \leq 4$ |  |          |                       |                                             |
|                       |               |                   |                      |  |          | %(OverP) <sup>a</sup> | % ( $\hat{\theta} \approx 0$ ) <sup>b</sup> |
| <i>All data</i>       |               |                   |                      |  |          |                       |                                             |
| EvE                   |               |                   |                      |  |          |                       |                                             |
| High                  | 3 (25%)       | 3 (25%)           | 6 (50%)              |  | 6   100% | 45                    | 37                                          |
| Low                   | 1 (8%)        | 5 (42%)           | 6 (50%)              |  | 2   33%  | 52*                   | 38*                                         |
| IBE                   |               |                   |                      |  |          |                       |                                             |
| High                  | 7 (58%)       | 0                 | 5 (42%)              |  | 1   20%  | 22                    | 12                                          |
| Low                   | 3 (25%)       | 0                 | 9 (75%)              |  | 1   11%  | 10                    | 6                                           |
| <i>Last 75 rounds</i> |               |                   |                      |  |          |                       |                                             |
| EvE                   |               |                   |                      |  |          |                       |                                             |
| High                  | 0             | 6 (50%)           | 6 (50%)              |  | 6   100% | 55                    | 43                                          |
| Low                   | 0             | 6 (50%)           | 6 (50%)              |  | 5   83%  | 56                    | 54                                          |
| IBE                   |               |                   |                      |  |          |                       |                                             |
| High                  | 4 (33%)       | 0                 | 8 (67%)              |  | 4   50%  | 27                    | 18                                          |
| Low                   | 2 (17%)       | 0                 | 10 (83%)             |  | 5   50%  | 27                    | 17                                          |

There is a total of 12 sessions per payoff level;  $K_{Min} = 1$  characterises homogenous players and  $1 < K_{Min} \leq 4$  cluster-heterogeneity; Detailed statistics refer to non-rejected specifications

<sup>a</sup>% of Over-Parametrised specifications with  $1 < K_{Min} \leq 4$

<sup>b</sup>% of insignificant/inconsistent estimates

<sup>c</sup>% of individuals with insignificant estimates

\*Including/relating to two inconsistent EvE estimates

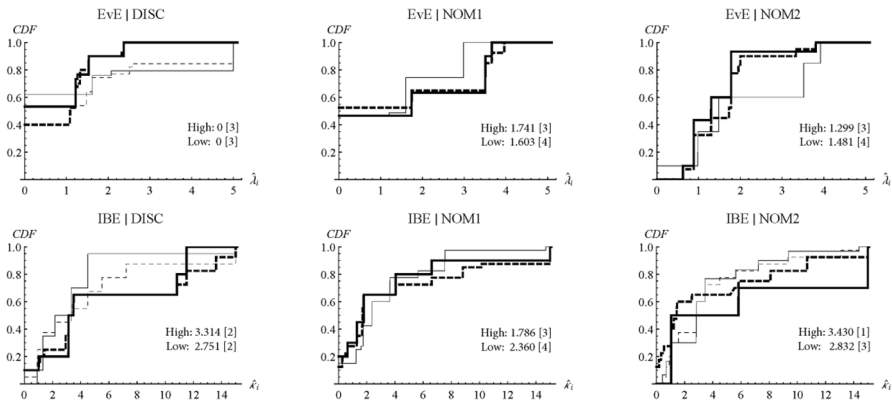
IBE-specifications do so. The summary tables in Appendix VII further reveal that both models are rejected for 4 sessions and that both are not rejected for 14 others. Since the remaining 6 sessions (25%) reject only IBE, it appears that EvE organises best the observed behaviour.

We proceed with checking whether the cluster-estimates of a specification (session) are heterogeneous with pairwise  $\chi^2$ -tests of equality and note that when all pairwise-tests are rejected, the estimates are considered heterogeneous if all pairwise-tests are also rejected when assuming  $K + 1$  clusters and the clusters were nested – the pairwise test outcomes are summarised in the last columns of Tables VII.A.1–4 in Appendix VII.A. On the other hand, a single non-rejection of equality implies that the specification is over-parametrised so the estimated cluster parameters are unreliable and one can only conclude that it has at most  $K_{Min} - 1$  clusters.

The last three columns of Table 3 refer to the non-rejected specifications of a treatment and report the percentages of (1) over-parametrised multi-clustered specifications, (2) insignificant or inconsistent estimates and (3) individuals affected by such estimates. The models sharply differ according to these criteria as EvE's specifications are far more likely to be over-parametrised than the IBE ones no matter the payoff level, i.e., a five-fold (three-fold) percentage difference when payoffs are High (Low). Most estimates of non-over-parametrised EvE-specifications are insignificant and none of these specifications yields estimates that fulfil the conditions to be considered heterogeneous. As for IBE, all estimates of non-over-parametrised specifications comply with these conditions when  $1 < K_{Min} \leq 3$  which leads us to conclude that, as expected, IBE accommodates cluster-heterogeneity better than EVE (cf. Sect. 4). Finally, about 50% of EvE's estimates are insignificant and affect some 37% of individuals no matter the payoff level whereas for IBE the figures drop at least by half, especially when payoffs are Low.

We highlight treatment differences by assigning to each participant the  $\theta$ -estimate of the cluster s/he belongs to and by comparing the resulting cumulative distributions of estimates for High and Low payoffs in each payoff structure. These distributions are displayed in Fig. 5 (with the samples' median estimates)—insignificant estimates were set equal to 0. To document the effect of the  $\Sigma$ -test on inference, the plots assume either (1) all estimates regardless of the sessions'  $\Sigma$ -test outcomes (cf. dashed lines), or (2) estimates of non-rejected specifications only (cf. plain lines). In this regard, the distributions pertaining to (1) and (2) reveal important differences only when non-rejected specifications are seldom, as for IBE in NOM2/High.

The distributions' large steps witness the presence of prominent clusters. In the case of EvE, the most prominent clusters consist of insignificant estimates and are found in DISC and NOM1 no matter the payoff level. There are also noticeable clusters of relatively large estimates supporting a more intense exploitation when payoffs are Low in DISC (with  $\hat{\lambda}_i \geq 5$  for over 20% of participants) and in NOM2 (with  $\hat{\lambda}_i \geq 3.5$  for about 40%). Such larger estimates counter-intuitively suggest that for



**Fig. 5** Cumulative distributions of individuals' OLS estimates. Thick (Thin) lines stand for High (Low) payoff levels—dashed lines refer to the  $4 \times 10$  estimates of a treatment regardless of the  $\Sigma$ -test outcomes. Insignificant estimates are set equal to 0. The plots report the estimates medians and numbers of non-rejected specifications (in brackets). The CDFs assume a maximum  $\hat{\lambda}_i$ - and  $\hat{\kappa}_i$ -estimates of 5 and 15, respectively

these participants, exploitation is more intense when payoffs are Low. This contrasts with NOM1 where the distributions are more alike across payoff levels and support the prediction that exploitation intensifies with payoffs, as the median estimates also qualitatively suggest.

As for IBE, the distributions look similar in NOM1 and suggest no particular 'pay-off magnitude' effect. They also display no prominent clusters of large estimates and thus contrast with the distributions of DISC and NOM2 which both do when payoffs are High ( $\hat{\kappa}_i > 10$  for about 30% of participants in these treatments). The presence of such clusters in those treatments identifies participants with low entry-rates, and their absence in NOM1 is in line with Observation 1(B): (1) the distributions and median estimates suggest that  $\kappa_{High} > \kappa_{Low}$  in DISC and NOM2, and  $\kappa_{High} \approx \kappa_{Low}$  in NOM1, and (2) the median estimates support  $\kappa_{NOM1} < \kappa_{DISC} \approx \kappa_{NOM2}$  when payoffs are High.

We attribute the absence of such clusters in NOM1 and the higher participation in NOM1/High to the relatively lower risk of regretting to enter that this structure entails when compared to DISC (which yields zero payoffs in case of over-entry) or to NOM2 (which bears an incentive to enter to avoid the risk of under-entry but which highest payoffs obtain only when  $A = \{3, 4, 5\}$ , cf. Appendix IV.D).

All in all, allowing for cluster-heterogeneity in the estimations reveals important differences in the models' explanatory powers and indicates that IBE outperforms EvE in this regard. We summarize this as follows:

**Observation 2:** When relaxing symmetry and estimating the models with OLS procedures, the null of the  $\Sigma$ -test is less likely to be rejected by EvE than by IBE (17% vs 42% of all sessions, respectively). However, when compared to IBE, the non-rejected EvE-specifications are: (1) less likely to reject homogeneity, (2) more likely to be over-parametrised, (3) more likely to generate insignificant or inconsistent estimates that affect a larger proportion of participants,

and (4) unable to rationalise the presence of clusters of players with low entry-probabilities. Thus IBE accommodates cluster-heterogeneity better than EvE.

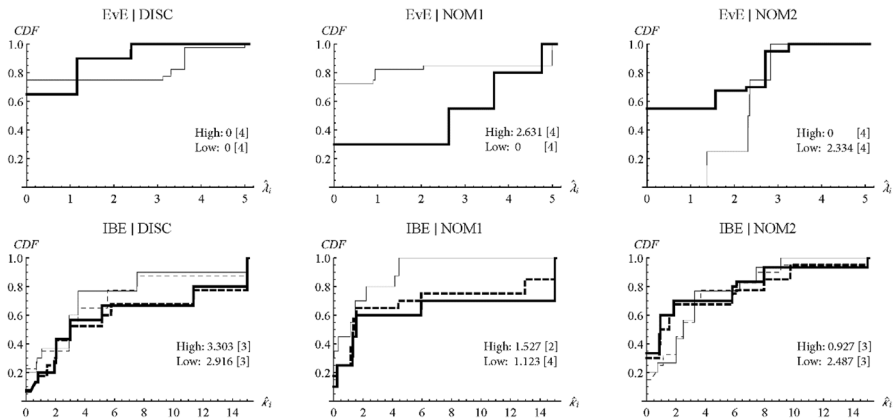
We conduct the same analysis for the last 75 rounds to check for a possible experience effect in the observed behaviour. The tests' outcomes are summarised in the lower panel of Table 3—see Tables VII.B.1–4 in Appendix VII.B for detailed results.<sup>17</sup> Now EvE is not rejected for *all* sessions whereas IBE is not rejected for 18 of them (75%, instead of 58% when accounting for all rounds) mostly with Low payoffs. Homogeneity ( $K_{Min} = 1$ ) is again rejected for IBE in all sessions, and it is not for EvE in 12 sessions so that 50% of EvE-specifications are multi-clustered (instead of 60%). These specifications also display fewer clusters only when assuming EvE in DISC and NOM1/High so behaviour in these treatments would become more homogenous in the long run according to EvE. Overall, since both models are not rejected for 18 (75%) sessions and the remaining 6 reject IBE but not EvE (cf. Appendix VII.B), EvE would appear again to organise the observed behaviour best.

Looking into the specifications' details, we find that the models yield more non-rejected over-parametrised specifications: over 83% for EvE, and 50% for IBE no matter the payoff level. There is also no evidence of heterogeneous estimates in the unique non-over-fitted EvE-specification (cf. NOM1/Low/Session 1) whereas all IBE-specifications with  $1 < K_{Min} \leq 3$  are heterogeneous. The models' differences remain in terms of insignificant estimates, with 55% of EvE-estimates indicating maximal exploration and affecting about 50% of participants whilst only 27% of the IBE ones are insignificant and concern 17% of participants no matter the payoff level.

The distributions of estimates in Fig. 6 tend to confirm the patterns found when assuming all data and they are moderately affected by the data-attrition resulting from the  $\Sigma$ -test rejections. Insignificant EvE-estimates are frequent in all treatments but NOM2/Low, where the estimates support a contained exploitation and the null of homogeneity ( $K_{Min} = 1$ ) in all sessions. Otherwise, the distributions pertaining to DISC and NOM2 still counter-intuitively suggest that it increases when payoffs are Low whereas those of NOM1 comply with the alternative that exploitation increases with payoff levels.

For IBE, the distributions and median estimates of DISC look alike those in Fig. 5 whilst those of NOM1 and NOM2 reveal (1) a drop in the median estimate of NOM1/Low and the presence of a cluster with large  $\kappa_i$ -estimates in NOM1/High, and (2) the absence of such a cluster in NOM2/High. However, the evolution of play in session data suggests that such higher (lower) participation for some participants in NOM1/Low and NOM2/High (NOM1/Low) are actually due to an 'end-game effect' in the last 10–20 rounds of these treatments, cf. Appendix V. This leads to the following observation:

<sup>17</sup> We also checked for significant trends in the estimates for batches of 75, 50 or 30 rounds that would reflect some learning but found no compelling evidence of declining trends in the IBE-estimates (suggesting a dampening loss aversion) or of increasing trends in the EvE-estimates (suggesting a reduced exploration over time) to report.



**Fig. 6** Cumulative distributions of individuals' OLS estimates (last 75 rounds). Thick (Thin) lines stand for High (Low) payoff levels—dashed lines refer to the  $4 \times 10$  estimates of a treatment regardless of the  $\Sigma$ -test outcomes. Insignificant estimates (at  $\alpha = 5\%$ ) are set equal to 0. The plots report the estimates medians and numbers of non-rejected specifications (in brackets). The CDFs assume a maximum  $\hat{\lambda}_T$ - and  $\hat{\kappa}_T$ -estimates of 5 and 15, respectively

Observation 3: When relaxing symmetry and estimating the models with OLS procedures and the data of the last 75 rounds:

- (A) The null of the  $\Sigma$ -test is less likely to be rejected by EvE than by IBE (0% of all sessions vs 25% for IBE).
- (B) Non-rejected EvE-specifications in DISC/High and especially NOM1/High have fewer clusters so behaviour becomes more homogenous in the long run according to EvE.
- (C) The features 1) to 4) of EvE's non-rejected specifications outlined in Observation 2 hold and confirm IBE's superior ability in organising the observed behaviour.

We proceed with a second robustness check of Observation 2 by estimating the models with more efficient procedures that possibly call for ('naïve' or Tikhonov) regularisation of the error variance matrix to address the unstable results we got when estimating the models with GLS methods. We thus consider five minimum-distance estimators in addition to the OLS and GLS ones, and we allow for regularisation whenever it is deemed necessary to give the models their best shot at organising the data.<sup>18</sup> That is, we estimated the models and their inverse forms for each session with seven estimators and with  $K = \{1, 2, 3, 4\}$  clusters, generating over 80 specifications per model and treatment. For each model and value of  $K$ , we selected the specification that best addresses a set of criteria regarding the theoretical consistency

<sup>18</sup> Without regularisation, we found exaggeratedly small estimated variances in the case of EvE and counter-intuitive, negative, estimates in the case of IBE—such estimates may happen when using Feasible Generalized Least Squares with a non-diagonal weighting matrix, as was the case. The details of the regularisation procedures and the 'minimum distance' estimators that we used are relegated to Appendix III.D.

**Table 4** Summary of specification test outcomes with(out) regularisation

|            | #(Rejections) | #(Non-rejections) |                |                        |                                             |                         |
|------------|---------------|-------------------|----------------|------------------------|---------------------------------------------|-------------------------|
|            |               | $K = 1$           | $1 < K \leq 4$ | Details                |                                             |                         |
|            |               |                   |                | % (OverP) <sup>a</sup> | % ( $\hat{\theta} \approx 0$ ) <sup>b</sup> | % (Indiv.) <sup>c</sup> |
| <hr/>      |               |                   |                |                        |                                             |                         |
| <i>EvE</i> |               |                   |                |                        |                                             |                         |
| High       | 5 (42%)       | 1 (8%)            | 6 (50%)        | 3   50%                | 11                                          | 13                      |
| Low        | 1 (8%)        | 1 (8%)            | 10 (84%)       | 8   80%                | 18*                                         | 19*                     |
| <hr/>      |               |                   |                |                        |                                             |                         |
| <i>IBE</i> |               |                   |                |                        |                                             |                         |
| High       | 7 (58%)       | 0                 | 5 (42%)        | 1   20%                | 17                                          | 10                      |
| Low        | 3 (25%)       | 0                 | 9 (75%)        | 0                      | 10                                          | 6                       |

There is a total of 12 sessions per payoff level;  $K_{Min} = 1$  characterises homogenous players and  $1 < K_{Min} \leq 4$  cluster-heterogeneity; Detailed statistics refer to non-rejected specifications

<sup>a</sup>% of over-parametrised specifications with  $1 < K_{Min} \leq 4$

<sup>b</sup>% of insignificant/inconsistent estimates

<sup>c</sup>% of individuals with insignificant estimates

\*Including/relating to two inconsistent EvE estimates

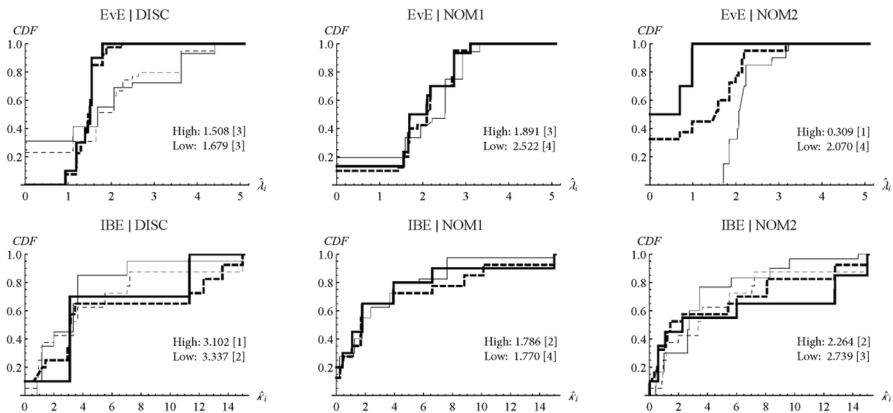
of parameter estimates and the credibility of their confidence intervals, but also to the condition number of the variance matrix of the error terms (not too large) and to the magnitude of the efficiency gains relative to OLS (not too large but not negligible).

The selected  $K_{Min}$ -specifications are reported in Tables I to IV of Appendix VII.C and indicate that some form of regularisation is needed for 19 sessions (79%) when estimating EvE and for only 2 (8%) when estimating IBE.<sup>19</sup> The  $\Sigma$ -test outcomes and the main characteristics of the models' non-rejected specifications are summarised in Table 4. They first indicate that the use of regularisation marginally affects the  $\Sigma$ -test outcomes for EvE and leaves those for IBE idle. Also, both models are not rejected for 13 sessions (instead of 14 when using OLS methods), EvE is not rejected for 6 (25%) and IBE for only 1 session (instead of 0).

The effect of regularisation is more salient on the estimates since the models become comparable in terms of rejecting homogeneity (i.e., 22 sessions for EvE vs 24 sessions for IBE), the proportion of insignificant/inconsistent estimates and, to a lesser extent, the proportion of individuals with such estimates. Yet, over 50% of EvE's non-rejected specifications are still over-parametrised whereas less than 20% of the IBE-ones are so.

The plots in Fig. 7 refer to heterogeneous samples of estimators and appear again to be affected by the  $\Sigma$ -test results only when the available data is sparse, as for EvE in NOM2/High. Insignificant EvE-estimates are mostly found in DISC/Low and NOM2/High, and they are about equally frequent no matter the payoff level in NOM1. Otherwise, the distributions of IBE-estimates, like those of EVE-estimates in DISC, display similar patterns as those referring to OLS estimates, cf. Fig. 5. The

<sup>19</sup> Most of the preferred EvE-specifications involve minimum distance estimators whereas the IBE-ones involve such estimators in only two instances and either standard OLS or GLS estimators otherwise.



**Fig. 7** Cumulative distributions of individuals' (regularized) estimates. Thick (Thin) lines stand for High (Low) payoff levels—dashed lines refer to the  $4 \times 10$  estimates of a treatment regardless of the  $\Sigma$ -test outcomes. Insignificant estimates are set equal to 0. The plots report the estimates medians and numbers of non-rejected specifications (in brackets). The CDFs assume a maximum  $\hat{\lambda}_i$ - and  $\hat{\kappa}_i$ -estimates of 5 and 15, respectively

most noticeable changes occur for the EVE-estimates of NOM1 and NOM2: they are now most similar across payoff levels in NOM1 and suggest no particular payoff magnitude effect (like the IBE-distributions of this treatment) whereas they are mostly different in NOM2, with stochastically larger (and mostly homogenous) cluster-estimates when payoffs are Low.

Overall, this robustness analysis confirms the models' respective (in)sensitivity to the symmetric assumption (Observation 1) and IBE's superior ability to diagnose a model-consistent cluster-heterogeneity in the observed behaviour (Observation 2). We summarise the above in the following final observation:

Observation 4: When relaxing symmetry and using (naïve or Tikhonov) regularisation procedures when estimating the models with GLS or distance-based estimators (instead of OLS estimators):

- (A) EvE is still less likely to reject the null of the  $\Sigma$ -test (25% of all sessions vs 42% for IBE).
- (B) The features 1) to 4) of EvE's non-rejected specifications outlined in Observation 2 hold and confirm IBE's superior ability in organising the observed behaviour.
- (C) Our conclusions for IBE are hardly affected by the use of regularisation procedures.

## 6 Conclusion

In this paper we propose a novel approach to the analysis of symmetric participation games that checks the consistency of a model's estimates with the restrictions it imposes on individual behaviour. This approach relaxes the model's assumption of



symmetry by allowing for the existence of clusters of players with similar observable characteristics, and it assesses how much cluster-heterogeneity a model can tolerate to still be consistent with its behavioural restrictions by means of a specification test. Thus, besides offering an alternative to the usual assessment of a model in terms of its goodness-of-fit, this approach allows for individual differences to be accounted for in a model-consistent way and therefore contributes to the literature on modelling heterogeneity in static games, see e.g., Rogers et al. (2009) and Golman (2011).<sup>20</sup>

We assessed this approach with data on market-entry experiments which we analyse in terms of two stationary models: Exploitation *versus* Exploration (EvE, which is equivalent Logit-QRE) and Impulse Balance Equilibrium (IBE). Our empirical analysis sheds new light on the models' sensitivities to the assumption of symmetric players or of cluster-heterogeneity and to the econometric procedures used. We summarise our findings in the following four points.

First, estimating EvE with the usual assumption of symmetric and homogeneous players provides limited insight into the analysis of behaviour in these games because (1) the session estimates are largely invariant to treatment conditions and mostly support a maximal exploration (or purely random behaviour), and (2) the estimates for the pooled data are seldom consistent with session estimates. In this regard, IBE outperforms EvE.

Second, when allowing for cluster-heterogeneity and estimating the models with OLS methods, the null of the specification test is less likely to be rejected for EvE, and EvE is more likely to support homogeneity than IBE. However, the estimated specifications have considerably more insignificant cluster-estimates and are typically over-parametrised, so IBE also outperforms EvE in terms of accommodating cluster-heterogeneity. This holds when the estimations pertain to the second half of the experiments to account for participants' experience of play.

Third, our approach can unveil behavioural patterns such as the presence of clusters of players with low-entry rates in some treatments and may explain them, i.e., such clusters are absent in treatments where payoffs remain positive when participation is over-capacity (as in NOM1) and they are present in treatments where the risk of experiencing a regret from entering is more salient (as in DISC and NOM2).

Fourth, when estimating the models with more efficient procedures (i.e., GLS or distance-based estimators that possibly allow for regularisation) our conclusions for IBE are hardly affected whereas those for EvE change considerably: homogeneity is then always rejected (like for IBE when assuming OLS methods) and insignificant or inconsistent cluster-estimates are less frequent. Yet, IBE still accommodates cluster-heterogeneity better than EvE.

<sup>20</sup> Another way to account for individual differences in symmetric games consists in reducing these games to decision-making problems which solutions encompass the Nash equilibrium as a special case, i.e., a specific parameter value. This approach discards the problem of modelling heterogeneity in strategic settings altogether and applies to games of incomplete information like auctions with private or common values, see Pezanis-Christou and Wu (2019a, 2019b).

Finally, the proposed approach is flexible enough to also allow an assessment of which type of heterogeneity is most consistent with some behavioural model, e.g., gender, socio-demographics, or any relevant mixture of observable characteristics. For example, it can be used to reveal a gender and/or a socio-demographic effect in the players' participation, and the specification test could determine whether this effect (or which of these effects) is consistent with the symmetric model considered.<sup>21</sup> It can also be applied to test predictions regarding the sorting of players into clusters of individuals who either always or never participate as a result of reinforcement learning, as Duffy and Hopkins (2005) predict and find. This, however, would raise the more challenging question of the formation of such clusters over time and its consistency with the type of learning considered. In this regard, our approach provides some first insights which we hope will be further explored.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10683-023-09797-8>.

**Acknowledgements** We thank Jacob Goeree, Ted Turocy, Tom Wilkening, and audiences at Aix-Marseille School of Economics (GREQAM), University of Melbourne, University of New South Wales, Queensland University of Technology, the International Meeting of the Economic Science Association (Hawaii), the Australia New-Zealand Workshop in Experimental Economics (Melbourne), the Workshop on Behavioral and Experimental Economics (Aix-en-Provence), the 11th Workshop on Complexity: Theory and Experimental Analyses (Nice), and the École des Hautes Études en Sciences Sociales (CAMS, Paris) for useful comments, as well as, in retrospect Reinhard Selten who suggested that we look into the Impulse Balance Equilibrium of this game. We also thank Nikolaos Georgantzis who generously gave us access to the Laboratory for Experimental Economics at Universidad Jaume I (Castellón, Spain) to run the experiments, Raquel Barreda and Juan Gómez for their assistance in their conduct, and Yohan Mathis and Mathilde Saguin for able research assistance on parts of the statistical analysis at the University of Strasbourg. We are grateful to the Editor, John Duffy, and two anonymous reviewers for valuable comments that greatly improved the presentation of our findings. Support from the European Commission through the *Laboratoires Européens Associés*, the Institute for Economic Analysis (Barcelona) and the University of Côte d'Azur (Nice, Project UCAinACTION of ANR-15-IDEX-01) where parts of this research were conducted, and the Australian Research Council (DP140102949) is gratefully acknowledged. The usual disclaimer applies.

**Funding** Open Access funding enabled and organized by CAUL and its Member Institutions.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

<sup>21</sup> See Babcock et al. (2017) who study the effect of gender in various volunteering type of dilemmas.

## References

- Anderson, S., de Palma, A., & Thisse, J.-F. (1992). *Discrete choice theory of product differentiation*. MIT Press.
- Anderson, S., & Engers, M. (2007). Participation games: Market entry, coordination and the beautiful blonde. *Journal of Economic Behavior and Organization*, 63, 120–137.
- Angeletos, G., & Lian, C. (2016). Incomplete information in macroeconomics: Accommodating frictions in coordination. NBER Working Papers 22297, National Bureau of Economic Research, Inc.
- Armantier, O., & Treich, N. (2009). Probability misperception in games: An application to the overbidding puzzle. *International Economic Review*, 50(4), 1079–1102.
- Arthur, W. (1994). Inductive reasoning and bounded rationality. *American Economic Review*, 84, 406–411.
- Babcock, L., Recalde, M., Vesterlund, L., & Weigart, L. (2017). Gender differences in accepting and receiving requests for tasks with low promotability. *American Economic Review*, 107(3), 714–747.
- Benartzi, S., & Thaler, R. H. (1995). Myopic loss aversion and the equity premium puzzle. *Quarterly Journal of Economics*, 110(1), 73–92.
- Bookstaber, R., & Kirman, A. (2018). Modelling a heterogeneous world. In Hommes, C., LeBaron, B. (Eds.), *Handbook of computational economics*, vol. 4, *heterogeneous agent modelling* (pp. 769–795). North Holland.
- Bottazzi, G., & Devetag, G. (2003). A laboratory experiment on the minority game. *Physica a: Statistical Mechanics and Its Applications*, 324, 124–132.
- Bouchaud, J.-P. (2013). Crises and collective socio-economic phenomena: Simple models and challenges. *Journal of Statistical Physics*, 151(3–4), 567–606.
- Branch, W., & McGough, B. (2018). Heterogeneous expectations and microfoundations in macroeconomics. In C. Hommes & B. LeBaron (Eds.), 4, *Handbook of computational economics*, vol. 4, *heterogeneous agent modelling* (pp. 3–59). North Holland.
- Brunner, C., Camerer, C., & Goeree, J. (2010). A correction and re-examination of «Stationary concepts for experimental 2x2 games». *American Economic Review*, 101, 1029–1040.
- Camerer, C., & Lovo, D. (1999). Overconfidence and excess entry: An experimental approach. *American Economic Review*, 89(1), 306–318.
- Camerer, C., Nunnari, S., & Palfrey, T. (2016). Quantal response and nonequilibrium beliefs explain overbidding in maximum-value auctions. *Games and Economic Behavior*, 98, 242–263.
- Challet, D., Marsili, M., & Zhang, Y. -C. (2014). *Minority games: Interacting agents in financial markets*. Oxford Finance Series.
- Costa-Gomes, M., Crawford, V., & Iriberri, N. (2009). Comparing models of strategic thinking in Van Huyck, Battalio, and Beil's coordination games. *Journal of the European Economic Association*, 7, 377–387.
- Crawford, V. (2013). Boundedly rational versus optimization-based models of strategic thinking and learning in games. *Journal of Economic Literature*, 51, 512–527.
- Duffy, J., & Hopkins, E. (2005). Learning, information and sorting in market entry games: Theory and evidence. *Games and Economic Behavior*, 51, 31–62.
- Erev, I., Ert, E., & Roth, A. (2010). A choice prediction competition for market entry games: An introduction. *Games*, 1, 117–136.
- Erev, I., & Rapoport, A. (1998). Coordination, “Magic,” and reinforcement learning in a market entry game. *Games and Economic Behavior*, 23, 146–175.
- Evans, B., & Prokopenko, M. (2021). A maximum entropy model of bounded rational decision-making with prior beliefs and market feedback. *Entropy*, 23, 669.
- Fischbacher, U. (2007). z-Tree: Zurich toolbox for readymade economic experiments. *Experimental Economics*, 10(2), 171–178.
- Gabaix, X. (2018). Behavioral inattention. In D. Bernheim, S. Della Vigna, & D. Laibson (Eds.), *Handbook of behavioral economics*. North Holland.
- Goeree, J., & Holt, C. (2005). An explanation of anomalous behavior in models of political participation. *American Political Science Review*, 99(2), 201–213.
- Goeree, J., Holt, C., & Palfrey, T. (2005). Regular quantal response equilibrium. *Experimental Economics*, 8, 347–367.
- Goeree, J., Holt, C., & Palfrey, T. (2016). *Quantal response equilibrium: A stochastic theory of games*. Princeton University Press.

- Golman, R. (2011). Quantal response equilibria with heterogeneous agents. *Journal of Economic Theory*, 146, 2013–2028.
- Gouriéroux, C., & Monfort, A. (1995). *Statistics and econometric models, volumes 1–2*. Cambridge University Press.
- Haile, P., Hortaçsu, A., & Kosenok, G. (2008). On the empirical content of quantal response equilibrium. *American Economic Review*, 98(1), 180–200.
- Kahneman, D. (1988). Experimental economics: A psychological perspective. In R. Tietz, W. Albers, & R. Selten (Eds.), *Bounded rational behavior in experimental games and markets* (pp. 11–18). Springer.
- Karp, L., Lee, I., & Mason, R. (2007). A global game with strategic substitutes and complements. *Games and Economic Behavior*, 60(1), 155–175.
- Kirman, A. (2006). Heterogeneity in economics. *Journal of Economic Interactions and Coordination*, 1, 89–117.
- Kirman, A. (2011). *Complex economics—Individual and collective rationality*. Routledge.
- Matějka, F., & McKay, A. (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, 105(1), 272–298.
- McKelvey, R., & Palfrey, T. (1995). Quantal response equilibria for normal form games. *Games and Economic Behavior*, 10, 6–38.
- McKelvey, R., Palfrey, T., & Weber, R. (2000). The effects of payoff magnitudes and heterogeneity on behavior in 2x2 games with unique mixed strategy equilibria. *Journal of Economic Behavior and Organization*, 42, 523–548.
- Melo, E., Pogorelskiy, K., & Shum, M. (2019). Testing the quantal response hypothesis. *International Economic Review*, 60(1), 53–74.
- Meyer, D., Van Huyck, J., Battalio, R., & Saving, T. (1992). History's role in coordinating decentralized allocation decisions. *Journal of Political Economy*, 100(2), 292–316.
- Nadal, J.-P., Weisbuch, G., Chenevez, O., & Kirman, A. (1998). A formal approach to market organization: Choice functions, mean field approximation and maximum entropy principle. In J. Lesourne & A. Orléan (Eds.), *Advances in self-organization and evolutionary economics* (pp. 149–159). Economica.
- Ochs, J. (1990). Coordination in market entry games. In D. Budescu, I. Erev, & R. Zwick (Eds.), *Games and human behavior: Essays in honor of Amnon Rapoport* (pp. 143–172). Lawrence Erlbaum Associates.
- Ockenfels, A., & Selten, R. (2005). Impulse balance equilibrium and feedback in first-price auctions. *Games and Economic Behavior*, 51(1), 155–170.
- Pezanis-Christou, P., Wu, H. (2019b). Rationalizing (non-)equilibrium bidding in maximum value auctions without beliefs about others' behaviour. *SSRN Working Paper #3233164*.
- Pezanis-Christou, P., & Wu, H. (2019a). An individual decision-making approach to bidding in first-price and all-pay auctions. *SSRN Working Paper #3301838*.
- Rapoport, A. (1995). Individual strategies in a market entry game. *Group Decision and Negotiations*, 4, 117–133.
- Rapoport, A., Seale, D., Erev, I., & Sundali, J. (1998). Equilibrium play in large group market entry games. *Management Science*, 44(1), 119–141.
- Rapoport, A., Seale, D., & Ordóñez, L. (2002). Weighted probabilities in tacit coordination under uncertainty: Theory and evidence from market entry games. *Journal of Risk and Uncertainty*, 25, 21–45.
- Rogers, B., Palfrey, T., & Camerer, C. (2009). Heterogenous quantal response equilibrium and cognitive hierarchies. *Journal of Economic Theory*, 144, 1440–1467.
- Scharfenaker, E., & Yang, J. (2020). Maximum entropy economics: Where do we stand? *European Physical Journal Special Topics*, 229, 1573–1575.
- Selten, R., Abbink, K., & Cox, R. (2005). Learning direction theory and the winner's curse. *Experimental Economics*, 8, 5–20.
- Selten, R., & Buchta, J. (1999). Experimental sealed bid first price auctions with directly observed bid functions. In D. Budescu, I. Erev, & R. Zwick (Eds.), *Games and human behavior: Essays in honor of Amnon Rapoport* (pp. 79–104). Mahwah: Lawrence Erlbaum Associates.
- Selten, R., & Chmura, T. (2008). Stationary concepts for experimental 2x2 games. *American Economic Review*, 98(3), 938–966.
- Selten, R., Chmura, T., & Goerg, S. (2010). A correction and re-examination of «Stationary concepts for experimental 2x2 games»: A reply. *American Economic Review*, 101, 1041–1044.

- Selten, R., Schreckenberg, M., Pitz, T., Chmura, T., & Kube, S. (2007). Commuters' route choice behaviour. *Games and Economic Behavior*, 58, 394–406.
- Sundali, J., Rapoport, A., & Seale, D. (1995). Coordination in market entry games with symmetric players. *Organizational Behavior and Human Decision Processes*, 64(2), 203–218.
- Tversky, A., & Kahneman, D. (1991). Loss aversion and riskless choice: A reference dependent model. *Quarterly Journal of Economics*, 107, 1039–1061.
- Weisbuch, G., Kirman, A., & Herreiner, D. (2000). Market organisation and trading relationships. *Economic Journal*, 110, 411–436.
- Weizsäcker, G. (2003). Ignoring the rationality of others: Evidence from experimental normal form games. *Games and Economic Behavior*, 44, 145–171.
- Yeung, C., & Zhang, Y. -C. (2009). Minority games. *Encyclopedia of Complexity and Systems Science* 5588–5604.
- Zwicky, R., & Rapoport, A. (2002). Tacit coordination in a decentralized market entry game with fixed capacity. *Experimental Economics*, 5, 253–272.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.