ECARES

Equilibrium Play in First Price Auctions: Revealed Preference Analysis

Laurens Cherchye
Department of Economics, KU Leuven

Thomas Demuynck
SBS-EM, ECARES, Université libre de Bruxelles

Bram De Rock
SBS-EM, ECARES, Université libre de Bruxelles
Department of Economics, KU Leuven

Mikhail Freer
SBS-EM, ECARES, Université libre de Bruxelles

December, 2018

ECARES working paper 2018-36

ECARES
ULB - CP 114/04
50, F.D. Roosevelt Ave., B-1050 Brussels BELGIUM
www.ecares.org

Equilibrium play in first price auctions: Revealed preference analysis

Laurens Cherchye[*] Thomas Demuynck[†] Bram De Rock[‡]
Mikhail Freer[§]

December 3, 2018

Abstract

We provide a revealed preference characterization of equilibrium behavior in first price sealed bid auctions. This defines testable conditions for equilibrium play that are intrinsically nonparametric, meaning that they do not require a (non-verifiable) specification of the individual utility functions. We characterize equilibrium play for a sequence of observations on private values and bids for a given individual. In a first step, we assume that the distribution of bids in the population is fully known. In a second step, we relax this assumption and consider the more realistic case that the empirical analyst can only use a finite number of i.i.d. observations drawn from the population distribution. We demonstrate the empirical usefulness of our conditions through an illustrative application to an existing experimental data set of Neugebauer and Perote (2008). This application also shows the potential of our nonparametric characterization to study the behavioral phenomena learning and fatigue at the individual level.

Keywords: first price auctions, equilibrium play, revealed preference characterization, testable implications, experimental data.

[*]Department of economics, University of Leuven. E. Sabbelaan 53, B-8500 Kortrijk, Belgium. E-mail: laurens.cherchye@kuleuven.be. Laurens Cherchye gratefully acknowledges the European Research Council (ERC) for his Consolidator Grant 614221. Part of this research is financed by the Fund of Scientific Research Flanders (FWO-Vlaanderen).

[†]ECARES, Université Libre de Bruxelles. Avenue F. D. Roosevelt 50, CP 114, B-1050 Brussels, Belgium. E-mail: thomas.demuynck@ulb.ac.be. Thomas Demuynck acknowledges financial support by the Fonds de la Recherche Scientifique-FNRS under grant nr F.4516.18

[‡]ECARES, Université Libre de Bruxelles, and Department of Economics, University of Leuven (KU Leuven). Avenue F. D. Roosevelt 50, CP 114, B-1050 Brussels, Belgium. E-mail: bderock@ulb.ac.be. Bram De Rock gratefully acknowledges FWO and FNRS for their financial support.

[§]ECARES, Université Libre de Bruxelles. Avenue F. D. Roosevelt 50, CP 114, B-1050 Brussels, Belgium. E-mail: mfreer@ulb.ac.be.

1 Introduction

The theory of auctions is one of the most fruitful research areas in both theoretical and applied economics. The most standard and well known auction model is the first price sealed bid auction (FPA). In the FPA model, each participant anonymously sends a bid to the auctioneer. The agent with the highest bid wins the auction and receives the item in return for paying her bid. In this paper, we provide a revealed preference characterization for equilibrium behavior in the FPA model, which defines testable conditions for equilibrium play. A distinguishing and attractive feature of our revealed preference conditions is that they do not require a (non-verifiable) specification of the individual utility functions. The conditions are intrinsically nonparametric and, therefore, robust to specification bias.

We have two main results. Our first result is of a theoretical nature and structures the testable nature of equilibrium behavior in a FPA context. We assume a sequence of observations on both private values and bids for a given individual, and we assume that the distribution of bids in the population is fully known. We derive a revealed preference characterization that defines necessary and sufficient conditions for the existence of a (Bernouilli) utility function that rationalizes the observed bids. Particularly, the function represents the observed bid behavior as equilibrium behavior for the FPA model, meaning that the individual chooses the bid that maximizes her expected utility given the distribution of bids in the population.

For this setting, the FPA model has strong testable empirical implications. Admittedly, the assumption that we observe not only the bids but also the evaluated individual's private values may seem overly strong. In this respect, however, we also show that, when private values are not observed, then the FPA model as such is no longer testable: any observed bid behavior can be rationalized as equilibrium behavior. In this case, the assumption of equilibrium play can (only) be used to define (partially) identifying structure on the distribution of private values. Particularly, it is not the case that *any* distribution can rationalize the observed bid behavior and as such we can test specific structure of the distribution.

Our second main result is of a practical nature and instrumental to applying our first theoretical result to empirical data. It relaxes the assumption that the distribution of bids in the population is fully observed. Instead, we make the more realistic assumption that the empirical analyst can only use a finite number of i.i.d. observations drawn from this distribution. For this setting, we construct a statistical hypothesis test for consistency of observed bidding behavior with the FPA model.

Interestingly, our methodology can be applied at the level of individual auction players, which we demonstrate through an application to the experimental data set of Neugebauer and Perote (2008). As we will explain, this data set has a number of appealing features in view of illustrating our methodology, including the absence of feedback to players between auction rounds. Based on our statistical test, we conclude that there is strong empirical support favoring the equilibrium play hypothesis. Subsequently, we document patterns for each individual separately with respect to learning and fatigue. This provides the first fully nonparamtric empirical evidence at the individual level of these well-studied behavioral phenomena.

The rest of this paper unfolds as follows. Section 2 positions our contribution in the relevant literature. Section 3 sets the stage by introducing our theoretical set-up and notation. Section 4 presents our main theoretical result, which characterizes equilibrium play for a fully known population distribution of bids. Section 5 subsequently presents our statistical hypothesis test for the case in which the empirical analyst can only use a finite number of i.i.d. observations from this population distribution. Section 6 discusses our empirical application. Section 7 concludes.

2 Literature overview

This paper touches upon several vibrant literatures. First, there is a growing interest in the revealed preference analysis of equilibrium behavior. See, among many others, Brown and Matzkin (1996), Cherchye, Demuynck, and Rock (2011),Carvajal, Deb, Fenske, and Quah (2013), Echenique, Lee, Shum, and Yenmez (2013) and Cherchye, Demuynck, De Rock, and Vermeulen (2017). A distinguishing feature of our current study is that we consider a game theoretic setting with incomplete information, while these papers dealt with the analysis of games characterized by complete information. Furthermore, we are the first paper that develops a revealed preference characterization of the FPA model.

There is also a sizeable experimental literature on auctions.[1] Cox, Roberson, and Smith (1982) and Cox, Smith, and Walker (1988) document deviations from risk-neutral Bayesian Nash equilibrium in a FPA context. Follow-up research has tried to explain this deviating behavior based on risk-aversion (Cox, Smith, and Walker, 1988, 1992; Chen and Plott, 1998; Andreoni, Che, and Kim, 2007), smooth optimization (Goeree, Holt, and Palfrey, 2002), learning (Ockenfels and Selten, 2005; Neugebauer and Selten, 2006), cognitive hierarchy models (Crawford and Iriberri, 2007) and behavioral phenomena like spite (Morgan, Steiglitz, and Reis, 2003) or regret (Filiz-Ozbay and Ozbay, 2007). The recent paper by Kirchkamp and Reiß (2011) finds that deviations from risk-neutral equilibrium behavior originate from failure to best-respond (assuming risk-neutral players) rather than from

[1]The following list of experimental research on auction theory is far from exhaustive. We refer to the recent survey of Kagel and Levin (2016) for a more comprehensive overview.

incorrect belief formation. However, the existing literature still does not present conclusive evidence on whether subjects exhibit equilibrium play or not.

Finally, there is a growing literature that focuses on the econometric analysis of auctions. See, for example, Guerre, Perrigne, and Vuong (2000) and Athey and Haile (2002, 2007). Most of this literature focuses on identifying the distribution of values from the observed distribution of bids. The present paper, by contrast, concentrates on testing whether the observed behavior is consistent with the FPA model, rather than on identifying the structure that underlies the observed bidding behavior (using behavioral consistency with equilibrium play as a key identifying assumption).

To some extent, the developed econometric techniques can also be applied to laboratory data (with known value distribution) in order to test whether subjects exhibit equilibrium play. For example, Bajari and Hortacsu (2005) compare the empirical performance of different structural decision models for some a priori specified parametric specification of the decision makers' utility functions. They first estimate the distribution of values imposed by some model, and subsequently compare this estimated distribution to the true (known) distribution of values.

From this perspective, the current paper presents an alternative approach to test the FPA model, which is based on revealed preference principles. Two distinguishing features are that (1) we can test for equilibrium behavior at the individual level rather than at the aggregate group level, and (2) our results do not depend on some (unverifiable) structure for the individual utility functions. Moreover, the existing methods to estimate value distributions generally assume that bidders are risk-averse (Zincenko, 2018), while we only require subjects to be expected utility maximizers (without imposing any structure on their risk attitude).[2]

3 Set-up and notation

We assume a first price auction (FPA) with $N = \{1, \ldots, n\}$ the set of participating players. Before the auction starts every player $i \in N$ receives a private value for the item, which we denote by $v_i \in [0, \bar{v}]$. After receiving their private values, all players simultaneously submit a bid b_i for the item. The player with the highest bid wins and pays her bid b_i in return for the auctioned item. The payoff for player i when winning the auction equals

$$x_i = v_i - b_i.$$

A player who does not win the auction receives a payoff of zero.

[2]The majority of existing papers estimates the decision makers' utility values. To retrieve the distribution of values, they require specific assumptions about the shape of the individual utility functions.

We assume an experimental setting with bids obtained from a random population of subjects (i.e. the participants of the experiment), who are (randomly) endowed with some private value for the auctioneered item. More formally, we assume a probability space of individuals $(\mathcal{J}, \mathcal{F}_{\mathcal{J}}, \nu)$, with $\mathcal{J}$ the set of individuals, $\mathcal{F}_{\mathcal{J}}$ a suitable σ-algebra on $\mathcal{J}$, and $\nu : \mathcal{F}_{\mathcal{J}} \to [0,1]$ a probability measure. Intuitively, participants in the auction are randomly drawn from the set $\mathcal{J}$ according to the measure ν. Further, we assume that there is a probability space of values $([0,\overline{v}], \mathcal{B}_{[0,\overline{v}]}, \mu)$ that determines how the participants in the auction receive their values for the auctioned item. We assume that $0 < \overline{v}$ and $\mathcal{B}_{[0,\overline{v}]}$ is the Borel σ-algebra on $[0,\overline{v}]$.

We denote by G the distribution of bids in the population. Bids are determined by some measurable random variable $B(j,v)$ on the product space $(\mathcal{J} \times [0,\overline{v}], \mathcal{F}_{\mathcal{J}} \otimes \mathcal{B}_{[0,\overline{v}]}, \nu \times \mu)$, where $\nu \times \mu$ is the product probability measure that is generated from independent draws from $(\mathcal{J}, \mathcal{F}_{\mathcal{J}}, \nu)$ and $([0,\overline{v}], \mathcal{B}_{[0,\overline{v}]}, \mu)$.[3] The value of $B(j,v)$ gives the (possibly random) bid for agent j when receiving value v in an n-player FPA. We define G to be the cumulative distribution function of this random variable $B(j,v)$,

$$G(b) = \int_{[0,\overline{v}]} \int_{\mathcal{J}} \mathbf{1}[B(j,v) \leq b] \, \nu(dj)\mu(dv),$$

with $\mathbf{1}[B(j,v) \leq b]$ the indicator function that equals one if $B(j,v) \leq b$ and zero otherwise. If we assume that bids in an auction are independently drawn from the distribution G, then the distribution of highest bids (for $n-1$ players) is given by

$$H(b) = G(b)^{n-1}.$$

We impose the following assumption on the distribution function G.

Assumption 1. *The cumulative distribution function G satisfies the following conditions:*

- *G is Lipschitz continuous on $[0,\overline{v}]$ (i.e. there exists a λ such that for all $y, z \in [0,\overline{v}]$, $|G(z) - G(y)| \leq \lambda |z-y|$).*
- *For all $\varepsilon > 0$ we have that $G(\varepsilon) > 0$.*
- *$G(0) = 0$.*

[3] Specifically, we have in mind an experimental set-up where the values that are assigned to the different subjects are independent from the identity of the subjects. This assumption can easily be relaxed and is not crucial for our main results to remain valid.

The condition of Lipschitz continuity requires that the distribution function does not increase too quickly.[4] The second condition imposes that the probability of observing a strictly positive (possibly arbitrarily small) bid is strictly positive. If (1) bids are always less than the received private value and (2) the probability density of μ is strictly positive on $]0, \overline{v}]$, then this assumption wil be satisfied. Finally, the last condition assumes zero probability to observe a bid that is less than zero.

In the standard FPA model it is assumed that agent i has a utility function

$$u_i : [0, \bar{v}] \to \mathbb{R}_+.$$

The agent knows the distribution of bids G, and chooses (for given private value v_i) the bid b_i that maximizes her expected utility. Formally, b_i solves

$$\max_{b_i \in \mathbb{R}_+} H(b_i) u(v_i - b_i), \tag{1}$$

where we normalized the utility associated with not winning the auction to zero, $u_i(0) = 0$.

In Section 4, we will assume that the empirical analyst observes a finite number of values and bids for a given individual, and knows the full distribution of bids in the population. For this setting, we derive the nonparametric revealed preference conditions for behavioral consistency with the FPA model. We will also show that, if the empirical analyst would not observe the individual's values, then any observed bid behavior could be rationalized, meaning that the FPA model as such is no longer (nonparametrically) testable. In Section 5, we relax the assumption that the distribution of bids is fully known, and we consider the more realistic setting in which the empirical analyst can only use an estimate of this distribution.

4 Characterizing equilibrium play

We assume that the empirical analyst observes an individual who participates multiple times in a randomly matched n-player auction. The different auctions faced by the individual differ only in terms of her private values, which are randomly drawn according to $([0, \overline{v}], \mathcal{B}_{[0,\overline{v}]}, \mu)$, possibly under the control of the experimental designer.

[4]In fact, we can relax the assumption that the function G is Lipschitz continuous, but then we can only guarantee that the utility function that rationalizes the data is upper-semicontinuous instead of continuous. Recall, however, that upper semi-continuity is still sufficient for the function to achieve its maximum on compact domains.

In the current section, we consider the case where the distribution G of bids in the population is fully observed.[5] Then, the data set for the given player i can be represented by the tuple

$$D = \{(v_i^t, b_i^t)_{t=1}^T, G\},$$

which consists of a finite number of private values v_i^t, the corresponding bids b_i^t, and the distribution G. Throughout the paper, we will analyze the behavior of player i separately, without imposing any assumption on the behavior of the other players. Therefore, we will drop the subscript i from now onwards.

We say that the data set D is rationalizable in terms of the FPA model (*FPA-rationalizable*) if there exists a utility function u that represents the observed bids b^t as solving problem (1). The player chooses each bid in order to maximize her expected utility over the binary lottery that delivers a payoff $v^t - b$ with probability $(G(b))^{n-1}$ and a payoff of zero otherwise. This yields the following definition.

Definition 1. *A data set $D = \{(v^t, b^t)_{t=1}^T, G\}$ is* **FPA-rationalizable**, *if there exists a continuous, non-decreasing utility function $u : \mathbb{R} \to \mathbb{R}$, such that $u(0) = 0$, u is locally non-satiated at zero and, for all observations $t = 1, \ldots, T$,*

$$b^t \in \operatorname*{argmax}_{b \in \mathbb{R}} H(b)u(v^t - b) \text{ where } H(b) = (G(b))^{n-1}.$$

As before, we set the utility of losing the auction equal to zero, $u(0) = 0$. Obviously, this implies no loss of generality. Further, local non-satiation at zero requires, for all $\varepsilon > 0$, that there exists $x \in]-\varepsilon, \varepsilon[$ such that $u(x) > 0$. Together with non-decreasingness of u, this is equivalent to requiring $u(x) > 0$ for all $x > 0$. This condition is necessary to obtain testable conditions associated with FPA-rationalizability. Indeed, without this condition, we could trivially rationalize any observed behavior by using the constant utility function $u(x) = 0$ for all x. Moreover, local non-satiation at zero is necessary for players to have an incentive to participate to the auction in the first place.

The following lemma states a first auxiliary result.[6]

Lemma 1. *If $D = \{(v^t, b^t)_{t=1}^T, G\}$ is FPA-rationalizable, and $v^t > 0$ for all $t = 1, \ldots, T$, then $b^t \in]0, v^t[$ and $H(b^t) > 0$.*

Thus, if $v^t > 0$ in each observation t, a data set D is FPA-rationalizable only if every bid b^t is strictly between zero and the value v^t, and the probability of winning the auction at every observed bid is strictly positive. The intuition is

[5]In principle, we could allow the distribution of bids G to depend on t as well. However, the more common experimental framework is the one in which subjects are repeatedly playing the auction with G fixed over observations, without feedback between trials.

[6]Appendix A contains the proofs of our results.

straightforward. If bids are either less than or equal to zero, higher than or equal to v^t, or if the probability of winning is zero, then the expected utility of the auction is less than or equal to zero. On the other hand, the agent can always get a strictly positive expected utility by bidding marginally below her private value.

We remark that the condition $v^t > 0$ in Lemma 1 is very weak in the sense that observing $v^t = 0$ for some t occurs with zero probability. In what follows, we will say that a data set D is *regular* if it satisfies the conditions in Lemma 1.

Definition 2. *The data set $D = \{(v^t, b^t)_{t=1}^T, G\}$ is* **regular** *if, for all $t = 1, \dots, T$, we have that $v^t > 0$, $b^t \in]0, v^t[$ and $H(b^t) > 0$.*

For every $t = 1, \dots, T$, we let $x^t = v^t - b^t$ represent the payoff associated with winning the auction in observation t. It follows directly from Lemma 1 that $x^t \in]0, v^t[$ for a regular data set. The following theorem provides the revealed preference conditions for a data set D to be FPA-rationalizable. Importantly, these conditions are intrinsically nonparametric as they do not require an explicit specification of the utility function u.

Theorem 1. *A regular data set $D = \{(v^t, b^t)_{t=1}^T, G\}$ is FPA-rationalizable if and only if, for all $t = 1, \dots, T$, there exist numbers $U^t > 0$ such that, for all $t, s = 1, \dots, T$,*

$$H(v^t - x^t)U^t \geq H(v^t - x^s)U^s. \tag{2}$$

Theorem 1 presents a set of inequalities that give necessary and sufficient conditions for FPA-rationalizability. These inequalities are linear in the unknown numbers U^t, which makes them easy to check. Intuitively, every number U^t represents the utility of winning the auction in period t, i.e. $U^t = u(x^t)$. Further, condition (2) corresponds to the individual's maximization problem in Definition 1. In particular, the expected utility of using the observed bid b^t should be at least as high as the expected utility of making any other bid, including the bid $v^t - x^s$. This yields the condition

$$\begin{aligned} H(v^t - x^t)U^t &= H(b^t)u(v^t - b^t), \\ &\geq H(v^t - x^s)u(v^t - v^t + x^s), \\ &= H(v^t - x^s)U^s. \end{aligned}$$

Observe that condition (2) only needs to be verified for t, s with $v^t - x^s > 0$, as otherwise the right hand side equals zero by definition, which makes that the inequality is satisfied automatically.

The necessity of condition (2) is relatively straightforward and may seem a rather weak implication. Interestingly, however, Theorem 1 shows that data consistency with condition (2) is not only necessary but also sufficient for FPA-rationalizabiliy. Particularly, in Appendix A.2 we provide a constructive proof that specifies a data rationalizing utility function based on the inequality condition (2).

Example 1 provides a practical illustration of the rationalizability condition in Theorem 1. It contains a data set that does not satisfy the condition. This shows that our nonparametric condition for FPA-rationalizability has substantial empirical bite as soon as the data set contains at least two observations.

Example 1. *Assume a data set D with two observations t, v such that*

$$H(v^t - x^t) = \frac{1}{10} \qquad H(v^t - x^s) = \frac{1}{4},$$
$$H(v^s - x^t) = \frac{1}{3} \qquad H(v^s - x^s) = \frac{1}{2}.$$

Then, condition (2) *in Theorem 1 requires that*

$$\frac{1}{10}U^t \geq \frac{1}{4}U^s \Leftrightarrow \frac{U^t}{U^s} \geq 2.5, \text{ and}$$
$$\frac{1}{2}U^s \geq \frac{1}{3}U^t \Leftrightarrow \frac{U^t}{U^s} \leq 1.5,$$

which is impossible. We conclude that the data set is not FPA-rationalizable.

One important final remark pertains to the fact that our characterization of equilibrium play in Theorem 1 assumes that the individual's private values are observed by the empirical analyst. This raises the question whether the FPA model has any testable implications if these private values were not observed. We will conclude this section by showing that the answer to this question is negative. Particularly, if only the bids b^t and the distribution G are observed, then we can always construct values $\tilde{v}^t$ such that the data set $\{(\tilde{v}^t, b^t)_{t=1}^T, G\}$ satisfies condition (2) in Theorem 1, which implies that the data set is FPA-rationalizable.

To formally demonstrate our non-testability result, we let

$$\Delta \in \left]0, \bar{v} - \max_{t \in \{1,\ldots,T\}} b^t\right[.$$

Lemma 1 shows that Δ exists under FPA-rationalizability. For every $t = 1, \ldots, T$, we then consider the values $\tilde{v}^t = b^t + \Delta$, which are contained in $]0, \overline{v}[$. This specification of the values ensures $x^t = \tilde{v}^t - b^t = \Delta$, i.e. x^t is the same for each observation t. Furthermore, for all $t, s = 1, \ldots, T$, we let $U^t = U^s = 1$. All this gives

$$H(\tilde{v}^t - x^t)U^t = H(\tilde{v}^t - x^s)U^s = H(b^t),$$

which implies that the FPA-rationalizability condition (2) in Theorem 1 is satisfied. We thus obtain the following corollary.

Corollary 1. *For every data set* $D = \{(b^t)_{t=1}^T, G\}$ *with* $G(b^t) > 0$ *and* $b^t \in]0, \overline{v}[$*, there exist values* $\tilde{v}^t$ *such that* $D' = \{(\tilde{v}^t, b^t)_{t=1}^T, H\}$ *is regular and FPA-rationalizable.*

In a sense, Corollary 1 reproduces the non-identification result of Guerre, Perrigne, and Vuong (2009) for our specific setting. At this point, however, it is important to emphasize that our characterization in Theorem 1 can still be used to define (partially) identifying structure on the distribution of values. Corollary 1 does not imply that it is impossible to partially identify the underlying values that rationalize the observed behavior in terms of the FPA model. Indeed, we cannot conclude that, for a given data set $\{(b^t)_{t=1}^T, G\})$, any possible sequence of values $v^1, \ldots, v^T$ will satisfy condition (2) in Theorem 1. Putting it differently, we can test specific structure on the values distribution; it is not the case that any distribution can rationalize the observed bid behavior. For instance, Example 1 shows that, for a given distribution of bids, some specifications of the individual's private values may effectively reject the rationalizability condition in Theorem 1.

5 Unobserved distribution of bids

So far we have assumed that the distribution of bids G is fully observed by the empirical analyst. In this section, we show how to use our characterization in Theorem 1 if this strong observational assumption no longer holds. Particularly, we consider the more realistic scenario that the analyst can (only) construct an estimate of the distribution G from a finite sample of players' observed bids. In such a case, we can construct a statistical hypothesis test for FPA-rationalizability as characterized in Theorem 1.

More formally, we assume an empirical auction setting with n players, with the evaluated individual playing each time against a random draw of $n-1$ other participants (who themselves are also given random private values). This obtains a random sample of $m = T(n-1)$ i.i.d. bids b_j drawn from G. Then, it is possible to construct an estimator of the distribution G by using the empirical distribution function

$$\mathbb{G}_m(b) = \frac{1}{m} \sum_{j=1}^{m} \mathbf{1}[b_j \leq b],$$

with associated small sample bias

$$\varepsilon_m(b) = \mathbb{G}_m(b) - G(b).$$

We recall that our characterization in Theorem 1 only requires us to evaluate the distribution H (and therefore G) at a finite number of values $v^t - x^s$, with $v^t - x^s > 0$ for every $t, s \in \{1, \ldots, T\}$ (which automatically implies $G(v^t - x^s) > 0$ for all t, s). Correspondingly, it suffices to construct a finite vector of errors ε_m, with entries

$$(\varepsilon_m)_{t,s} = \mathbb{G}_m(v^t - x^s) - G(v^t - x^s).$$[7]

The vector $m^{1/2}\varepsilon_m$ has an asymptotic distribution that is multivariate normal with mean zero and variance-covariance matrix Ω, where

$$(\Omega)_{(t',s'),(t,s)} = \begin{cases} G(v^{t'} - x^{s'})(1 - G(v^t - x^s)) \text{ if } v^{t'} - x^{s'} < v^t - x^s \\ G(v^t - x^s)(1 - G(v^{t'} - x^{s'})) \text{ if } v^t - x^s < v^{t'} - x^{s'} \end{cases}.$$

Thus, standard results yield

$$m\,\varepsilon'\Omega^{-1}\varepsilon \sim^a \chi^2(K),$$

where $\sim^a$ denotes convergence in distribution and K is the size of the vector ε.[8]

Of course, in practice we do not observe the matrix Ω. We can approximate it using the finite sample analogue $\widehat{\Omega}_m$, where

$$(\widehat{\Omega}_m)_{(t',s'),(t,s)} = \begin{cases} \mathbb{G}_m(v_i^{t'} - x_i^{s'})(1 - \mathbb{G}_m(v_i^t - x_i^s)) \text{ if } v^{t'} - x^{s'} < v^t - x^s, \\ \mathbb{G}_m(v^t - x^s)(1 - \mathbb{G}_m(v^{t'} - x^{s'})) \text{ if } v^t - x^s < v^{t'} - x^{s'}. \end{cases}$$

Because $\widehat{\Omega}_m$ is a consistent estimate of Ω, it follows that

$$m\,\varepsilon'(\widehat{\Omega}_m)^{-1}\varepsilon \sim^a \chi^2(K),$$

We use this last result as the basis for our asymptotic hypothesis test for FPA-rationalizability. Specifically, consider the null hypothesis

$$H_0: \text{ the data set } D = \{(v^t, b^t)_{t=1}^T, G\} \text{ is FPA-rationalizable,}$$

To empirically check this hypothesis, we can solve the following minimization problem.

OP.I

$$Z_m = \min_{e_{t,s}\in\mathbb{R}, \hat{G}\in[0,1]^K, U^t>0} m\,(e_{1,1}, e_{1,2}, \ldots)'(\widehat{\Omega}_m)^{-1}(e_{1,1}, e_{1,2}, \ldots),$$

$$\text{s.t. } \forall t, s \text{ with } v^t - x^s > 0: \;\; e_{t,s} = \mathbb{G}_m(v^t - x^s) - \hat{G}_{t,s}, \tag{3}$$

$$(\hat{G}_{t,t})^{n-1}U^t \geq (\hat{G}_{t,s})^{n-1}U^s \tag{4}$$

[7]For simplicitly, we assume that all values $v^t - x^s$ are different. Obviously, this does not affect the core of our argument.

[8]See, for example, Sepanski (1994).

If the hypothesis H_0 holds true, we must have

$$Z_m \leq m\,\varepsilon'(\widehat{\Omega}_m)^{-1}\varepsilon,$$

Let us denote by c_α the $(1-\alpha)\times 100$th percentile of the $\chi^2(K)$ distribution. Then, if H_0 holds, we must have

$$\lim_{m\to\infty} \Pr[Z_m > c_\alpha] \leq \lim_{m\to\infty} \Pr[m\,\varepsilon'(\widehat{\Omega}_m)^{-1}\varepsilon > c_\alpha] = \alpha.$$

Thus, we can construct an asymptotic test of H_0 by solving **OP.I** for the given data set, to subsequently verify whether its solution value exceeds c_α.

Two concluding remarks are in order. First, our empirical hypothesis test is conservative in nature when compared to the theoretical test (based on Theorem 1) that uses the true distribution of bids G. In this sense, our asymptotic test bears similarities to Varian (1985)'s statistical test for rational consumer behavior (based on revealed preference conditions) in the case of measurement error. However, there is one important difference. An intrinsic problem of Varian's test is that the variance of the error distribution cannot be estimated and, therefore, needs to be assumed. By contrast, the variance-covariance matrix used in our test can be estimated consistently.

Second, implementing our hypothesis test in principle requires solving the nonlinear minimization problem **OP.I**, which is computationally difficult because the constraints (4) are nonlinear in unknowns. Interestingly, however, we can convert this problem into a minimization problem with linear constraints and a quasi-convex objective function, which is easily solvable through standard algorithms for finding global minima.[9] To see this, we define $g_{t,s} = \ln(G_{t,s})$ and $\beta^t = \ln(U^t)$ for t, s with $v^t - x^s > 0$. Then, we can linearize the constraints (4) as

$$(n-1)g_{t,t} + \beta^t \geq (n-1)g_{t,s} + \beta^s,$$

while the constraints (3) yields

$$e_{t,s} = \mathbb{G}_m(v^t - x^s) - \exp(g_{t,s}).$$

Taken together, this gives rise to the following problem with only linear constraints and a quasi-convex objective function.

OP.II

$$\min_{g_{t,s}} m\,(e_{1,1}, e_{1,2}, \ldots)'(\hat{\Omega}_m)^{-1}(e_{1,1}, e_{1,2}, \ldots),$$
$$\text{s.t. } (n-1)g_{t,t} + \beta^t \geq (n-1)g_{t,s} + \beta^s,$$

where we defined $e_{t,s} = \mathbb{G}_m(v^t - x^s) - \exp(g_{t,s})$ for all t, s with $v^t - x^s > 0$.

[9] In particular the minimum can be found by using the subgradient projection method (see Kiwiel, 2001). Moreover, it can be well approximated using standard techniques for local minimization (e.g. trust region).

6 Illustrative application

We demonstrate the practical usefulness of our statistical revealed preference test for equilibrium play through an empirical application to the experimental data set of Neugebauer and Perote (2008). We begin by briefly explaining the experimental design and the set-up of our empirical analysis. This motivates that the assumptions underlying our statistical test are plausible for the given data. Subsequently, we will consider two types of empirical exercises. First, we evaluate the results of our statistical test both in terms of pass rates and discriminatory power. Next, we use our test results to study the behavioral phenomena of learning and fatigue at the level of individual decision makers.

Set-up. Neugebauer and Perote (2008) conducted their FPA experiment on a sample of 28 subjects. Every subject participated in 50 rounds of a 7-player FPA. At the beginning of every round, every subject received an i.i.d. value drawn from the uniform distribution on $[0, 100]$. Subjects got no feedback between the different auction rounds, and the data set contains the observed bidding behavior for each of the 28 subjects.[10] Interestingly, the absence of feedback between the subsequent auctions allows us to exclude various behavioral reasons for overbidding. In particular, there is no winner or loser regret. Moreover, there is no basis for updating subjects' beliefs about the distribution of bids in the population.

Our FPA model assumes that every subject acts against the population of bidding functions. We believe this is a plausible assumption for our data set given that feedback between rounds is absent, and because subjects do not know with whom they are matched. Moreover, the submitted bids can be assumed to be independent as there is no communication between the subjects and groups are randomly formed. This implies that we can reasonably use the observed bids to estimate the bidding distribution G. At this point, we remark that the observed bids do contain multiple observations from the same individuals. However, this does not violate the i.i.d. assumption as long as we assume the population of players to equal the set of participants of the experiment.[11]

[10]We actually only consider a subset of Neugebauer and Perote (2008)'s original data set. In their original experiment, these authors also conducted a second session with every subject participation in 50 additional auctions. However, in this session the subjects did receive feedback after every auction round. As the absence of feedback is important to motivate the validity of our statistical test, we have excluded these data from our application. Further, Neugebauer and Perote's original study also contained 28 additional subjects who played 50 rounds with feedback in a first session, and 50 periods without feedback in a second session. For the same reason as before, we chose not to include these subjects in our empirical application.

[11]If the population equals all potential participants of the experiment (i.e. also the individuals who did not participate), then the i.i.d. assumption might be violated. In this case, a solution would be to include only one (random) bid for every participant in the auction, which would

For our main analysis, we have split the data per subject into five blocks of 10 rounds each. Our motivation for this is threefold. A first reason relates to the power of our statistical test. The critical values of our test statistic are based on the percentiles of the χ^2 distribution with degrees of freedom proportional to the number of observations squared. As an implication, increasing the number of observations does not necessarily improve the power of our test. For example, by going from 10 to 20 observations we scale up the degrees of freedom from 100 to 400, which leads to a substantial increase in our critical values. Including 10 observations per person appears to define a good compromise between fit and power. Secondly, restricting the number of observations used in the statistical test is also useful from a computational point of view. Although the optimization problem **OP.II** is well behaved, the number of constraints is also proportional to the square of the number of observations. Finally, an attractive feature of splitting up the sample into blocks of 10 rounds is that it allows us to investigate dynamic patterns in the data. Particularly, as we will explain below, it allows us to study learning behavior and fatigue.

In our test of equilibrium play for a given individual i, we compute $\mathbb{G}$ as the empirical distribution function of all bids except from the bids of player i herself. This gives a total of 27×10 observed bids when using five blocks of 10 rounds per subject. As a sensitivity check, we have also conducted our statistical test for blocks consisting of 8 and 12 rounds. Our results in Appendix B show that our main test results are robust for these alternative data splittings.

Test results. Table 1 summarizes our test results for each block of 10 rounds. We consider significance levels (α) of 5% and 10%. The first row in every block presents the number and percentage of individuals for which we do not reject the null hypothesis of FPA-rationalizable behavior. In the second row of each block, we show 95% confidence intervals for the percentage of individuals passing our test (meaning that H_0 was not rejected). These confidence intervals are computed by using the Clopper–Pearson procedure.

Finally, the bottom row in every block reports on the power of our test, which gives the probability of rejecting H_0 in the case of (simulated) random bidding behavior. We compute power by using an adaptation of Bronars (1987)'s procedure, which was originally developed to evaluate the power of revealed preference tests for rational consumer behavior. Particularly, we simulate random behavior by generating 500 data sets of random bids (using a uniform distribution) between

bring the data generating process closer to the i.i.d. assumption. However, for our exercise this would mean that we can use only 27 observation to estimate G, which is obviously far too low to rely on asymptotic convergence results. As all 28 subjects participated in the same session and every participant has seen all possible opponents, we believe that equating the total population with the population of participants is not overly restrictive for our data set.

Rounds		$\alpha \geq .05$	$\alpha \geq .10$
$1 \leq t \leq 10$	Pass Rate	14 (50%)	14 (50%)
	95% conf. interval	31% – 69%	31% – 69%
	Random subjects	40%	37%
$11 \leq t \leq 20$	Pass Rate	26 (93%)	24 (86%)
	95% conf. interval	77% – 99%	67% – 96%
	Random subjects	45%	42%
$21 \leq t \leq 30$	Pass Rate	22 (79%)	22 (79%)
	95% conf. interval	59% – 92%	59% – 92%
	Random subjects	44%	42%
$31 \leq t \leq 40$	Pass Rate	20 (71%)	20 (71%)
	95% conf. interval	51% – 87%	51% – 87%
	Random subjects	55%	53%
$41 \leq t \leq 50$	Pass Rate	24 (86%)	24 (86%)
	95% conf. interval	67% – 96%	67% – 96%
	Random subjects	46%	43%

Table 1: Non-rejection rates for H_0 – significance levels of 5% and 10%

zero and the observed value. Next, we compute the fraction of these randomly generated data sets for which we can reject H_0 at the given significance level. The lower this fraction, the higher the power of our tests.

Some interesting conclusions can be drawn from Table 1. First, the pass rates for the "real" subjects in our sample are always above the pass rates for the "randomly generated" subjects. This suggests that the real subjects behave substantially more rational than the random subjects. Interestingly, when considering the 95% confidence intervals for our pass rates, we conclude that the difference is also statistically significant in most cases. This shows that our statistical test is sufficiently powerful to effectively discriminate between rational and random behavior.

The results in Table 1 use generic significance levels of 5% and 10%. In this respect, we recall from our discussion in Section 5 that our statistical test is a conservative one when compared to the theoretical test (based on Theorem 1) that uses the true distribution of bids. Therefore, in a following exercise we use "power-adjusted" significance levels. Particularly, we set the critical values used in our test such that only 10% and 5% of our randomly generated data sets pass the test (meaning that we cannot reject H_0). By construction, these critical values correspond to statistical tests with power equal to 90% and 95%, respectively.

The results of this additional exercise are given in Table 2. Comparing these results with the ones in Table 1 gives an indication of the conservative nature of

Rounds		Power= .9	Power= .95
$1 \leq t \leq 10$	Pass Rate	8 (29%)	6 (21%)
	95% conf. interval	13% – 49%	8% – 41%
	Random subjects	10%	5%
$11 \leq t \leq 20$	Pass Rate	12 (43%)	8 (29%)
	95% conf. interval	25% – 63%	13% – 49%
	Random subjects	10%	5%
$21 \leq t \leq 30$	Pass Rate	16 (57%)	16 (57%)
	95% conf. interval	37% – 76%	37% – 76%
	Random subjects	10%	5%
$31 \leq t \leq 40$	Pass Rate	16 (57%)	14 (50%)
	95% conf. interval	37% – 76%	31% – 69%
	Random subjects	10%	5%
$41 \leq t \leq 50$	Pass Rate	18 (64%)	16 (57%)
	95% conf. interval	44% – 81%	37% – 76%
	Random subjects	10%	5%

Table 2: Non-rejection rates for H_0 – power-adjusted significance levels

our statistical test. Particularly, from Table 1 we learn that the use of generic significance levels (of 5% and 10%) still tolerates a substantial amount of the randomly generated data sets: we do not reject H_0 for about 40-50% of our random data sets. Table 2 shows that the pass rates drop considerably (i.e. by about 20%) if we adjust the significance levels to control for power. Interestingly, however, a significant fraction of the subjects in our sample do remain FPA-rationalizable. Moreover, the pass rates for actual behavior are again significantly above the pass rates for random behavior.

Learning and fatigue. We conclude our empirical exercise by studying the behavioral phenomena of learning and fatigue effects. The underlying intuition is that subjects may need some time to learn their optimal bidding behavior but, after a significant amount of repetitions, they may also get tired (or bored) and start to act less rational. Hence, subjects can be inconsistent with FPA-rationalizability in the initial time blocks, become consistent later (learning), and finally again exhibit inconsistent behavior (fatigue). In our following evaluation, we will exploit the fact that our rationalizability test can be performed at the level of the individual decision makers. As such, we are not restricted to only consider group dynamics. We can also evaluate how behavioral dynamics affect the individual consistency patterns.

Table 2 indicates a pattern of learning at the aggregate sample level. Generally, the difference in pass rates between actual and random behavior is more pronounced for the later blocks of 10 rounds. This provides stronger evidence for FPA-rationalizability in the later auction rounds. However, when considering the results in more detail, we see that the pass rates actually stabilize between the third and fifth time blocks. For example, in the last column of Table 2 the pass rate decreases slightly from the third time block to the fourth time block, and picks up again in the fifth time block. Interestingly, these trends fall in line with the ones reported by Neugebauer and Perote (2008) for the same data.

In order to better understand these aggregate patterns, we next take a closer look at the individual pass rates. Table 3 describes the four different patterns of behavior that we found in our data. From this table, we observe that most subjects in the sample satisfy the second or third scenario, which are all directly connected to learning effects. We also find evidence suggesting fatigue effects for some subjects, which explains that the aggregate pass rates in Table 2 are not strictly increasing, but rather stabilizing in the later blocks. More generally, these results show the usefulness of our revealed preference methodology to study dynamics of rationalizability at the level of individual decision makers. This can provide useful insights to explain observed dynamics at the aggregate level.

Patterns	Power= .9	Power= .95
No pattern at all: H_0 is either rejected or not rejected for each of the five blocks	2	4
Learning pattern: H_0 is rejected for the earlier blocks but not for the later blocks	14	10
Learning and fatigue pattern: H_0 is rejected for the earlier blocks, followed by non-rejection in the middle blocks and rejection in the last blocks	6	6
Opposite pattern: H_0 not rejected in the earlier blocks, rejected in the middle blocks, and not rejected in the last blocks	6	8

Table 3: Different patterns for learning and fatigue

7 Concluding discussion

We provided a revealed preference characterization of equilibrium play in the FPA model. Our characterization is intrinsically nonparametric, which makes it robust to specification bias. We derived testable necessary and sufficient conditions such that observed bidding behavior is consistent with expected utility maximization for a known distribution of bids in the population. Building on this characterization, we next developed a statistical test that can be used in empirical settings, where the population distribution of bids is not fully observed.

We also demonstrated the practical usefulness of this statistical test through an application to experimental data. We could not reject the hypothesis of equilibrium for a large fraction of the subjects in the experiment. At a more general level, our results suggest that observed bidding behavior is largely consistent with equilibrium play as long as we control for behavioral biases and learning and relax implausible assumptions about players' preferences (such as risk-neutrality). Next, our analysis showed the value added of our methodology in terms of analyzing equilibrium play at the level of individual decision makers (e.g. to investigate learning and fatigue effects).

We see multiple directions for follow-up research. Most notably, our nonparametric characterization assumed that individuals are expected utility maximizers, without imposing any structure on their risk attitude. Future research can focus on extending our characterization by relaxing or strengthening our assumptions regarding individual preferences. For example, we may consider the weaker assumption that individual utility functions are continuous and satisfy first-order stochastic dominance. More formally, this means that an individual's utility is increasing in both the payoff x and the probability of winning $H(v - x)$. This extension would lead to a test of the FPA model for preferences that (only) satisfy first order stochastic dominance.

Conversely, we may also want to extend our framework by incorporating stronger restrictions on preferences than the ones we imposed in the current paper. For example, we may assume that agents are expected utility maximizers with risk-averse utility functions (implying concave utility functions). In this respect, an immediate consequence of risk-aversion is that the bidding function (i.e. the optimal bid as a function of the value) must be increasing. Interestingly, when checking this implication for the data set we used in our empirical application, we find that the implication holds for only about 50% of the subjects for whom we do not reject FPA-rationalizability. In other words, a substantial fraction of FPA-rationalizable subjects do not have a monotone bidding functions and, therefore, cannot be risk-averse.

References

James Andreoni, Yeon-Koo Che, and Jinwoo Kim. Asymmetric information about rivals' types in standard auctions: An experiment. *Games and Economic Behavior*, 59(2):240–259, 2007.

Susan Athey and Philip A Haile. Identification of standard auction models. *Econometrica*, 70:2107–2140, 2002.

Susan Athey and Philip A Haile. Nonparametric approaches to auctions. *Handbook of econometrics*, 6:3847–3965, 2007.

Patrick Bajari and Ali Hortacsu. Are structural estimates of auction models reasonable? evidence from experimental data. *Journal of Political Economy*, 113 (4):703–741, 2005.

Stephen G Bronars. The power of nonparametric tests of preference maximization. *Econometrica*, 55:693–698, 1987.

Donald J Brown and Rosa L Matzkin. Testable restrictions on the equilibrium manifold. *Econometrica*, 64:1249–1262, 1996.

Andres Carvajal, Rahul Deb, James Fenske, and John K-H Quah. Revealed preference tests of the cournot model. *Econometrica*, 81(6):2351–2379, 2013.

Kay-Yut Chen and Charles R Plott. Nonlinear behavior in sealed bid first price auctions. *Games and Economic Behavior*, 25(1):34–78, 1998.

Laurens Cherchye, Thomas Demuynck, and Bram De Rock. Revealed preference analysis of non-cooperative household consumption. *The Economic Journal*, 121 (555):1073–1096, 2011.

Laurens Cherchye, Thomas Demuynck, Bram De Rock, and Frederic Vermeulen. Household consumption when the marriage is stable. *American Economic Review*, 107(6):1507–34, 2017.

James C Cox, Bruce Roberson, and Vernon L Smith. Theory and behavior of single object auctions. *Research in experimental economics*, 2(1):1–43, 1982.

James C Cox, Vernon L Smith, and James M Walker. Theory and individual behavior of first-price auctions. *Journal of Risk and Uncertainty*, 1(1):61–99, 1988.

James C Cox, Vernon L Smith, and James M Walker. Theory and misbehavior of first-price auctions: Comment. *The American Economic Review*, 82(5):1392–1412, 1992.

Vincent P Crawford and Nagore Iriberri. Level-k auctions: Can a nonequilibrium model of strategic thinking explain the winner's curse and overbidding in private-value auctions? *Econometrica*, 75(6):1721–1770, 2007.

Federico Echenique, Sangmok Lee, Matthew Shum, and M Bumin Yenmez. The revealed preference theory of stable and extremal stable matchings. *Econometrica*, 81(1):153–171, 2013.

E Filiz-Ozbay and Erkut Y Ozbay. Auctions with anticipated regret: Theory and experiment. *The American Economic Review*, 97(4):1407–1418, 2007.

Jacob K Goeree, Charles A Holt, and Thomas R Palfrey. Quantal response equilibrium and overbidding in private-value auctions. *Journal of Economic Theory*, 104(1):247–272, 2002.

Emmanuel Guerre, Isabelle Perrigne, and Quang Vuong. Optimal nonparametric estimation of first-price auctions. *Econometrica*, 68:525–574, 2000.

Emmanuel Guerre, Isabelle Perrigne, and Quang Vuong. Nonparametric identification of risk aversion in first-price auctions under exclusion restrictions. *Econometrica*, 77(4):1193–1227, 2009.

John Kagel and Dan Levin. Auctions: A survey of experimental research. In John H Kagel and Alvin E Roth, editors, *The handbook of experimental economics, volume 2*, chapter 9, pages 563–638. Princeton university press, 2016.

Oliver Kirchkamp and J Philipp Reiß. Out-of-equilibrium bids in first-price auctions: Wrong expectations or wrong bids. *The Economic Journal*, 121(557): 1361–1397, 2011.

Krzysztof C Kiwiel. Convergence and efficiency of subgradient methods for quasiconvex minimization. *Mathematical programming*, 90(1):1–25, 2001.

John Morgan, Ken Steiglitz, and George Reis. The spite motive and equilibrium behavior in auctions. *Contributions in Economic Analysis & Policy*, 2(1), 2003.

Tibor Neugebauer and Javier Perote. Bidding 'as if' risk neutral in experimental first price auctions without information feedback. *Experimental Economics*, 11 (2):190–202, 2008.

Tibor Neugebauer and Reinhard Selten. Individual behavior of first-price auctions: The importance of information feedback in computerized experimental markets. *Games and Economic Behavior*, 54(1):183–204, 2006.

Axel Ockenfels and Reinhard Selten. Impulse balance equilibrium and feedback in first price auctions. *Games and Economic Behavior*, 51(1):155–170, 2005.

Steven J Sepanski. Asymptotics for multivariate t-statistic and hotelling's t2-statistic under infinite second moments via bootstrapping. *Journal of Multivariate Analysis*, 49(1):41–54, 1994.

Hal R Varian. Non-parametric analysis of optimizing behavior with measurement error. *Journal of Econometrics*, 30(1-2):445–458, 1985.

Federico Zincenko. Nonparametric estimation of first-price auctions with risk-averse bidders. *Journal of Econometrics*, 205:303–335, 2018.

A Proofs

A.1 Proof of Lemma 1

Let $v^t > 0$. If $b^t \leq 0$, then we know that $G(b^t) = 0$ and $H(b^t) = 0$. Also, if $b^t \geq v^t$ then $u(v^t - b^t) \leq 0$. This implies in both cases that

$$H(b^t)u(v^t - b^t) \leq 0.$$

Given our assumptions related to G, we have that $v^t > 0$ implies $G(v^t) > 0$. Therefore there exists a $\varepsilon > 0$, small enough, such that, $G(v^t - \varepsilon) > 0$ and $H(v^t - \varepsilon) > 0$. Then, if the agent chooses the bid $v^t - \varepsilon$, we have

$$H(v^t - \varepsilon)u(\varepsilon) > 0 \geq H(b^t)u(v^t - b^t).$$

This implies that a bid $b^t \leq 0$ or $b^t \geq v^t$ can never maximize (1), which proves this lemma.

A.2 Proof of Theorem 1

In the main text we showed the necessity of the conditions. To prove the sufficiency, we will construct a utility function $u : \mathbb{R} \to \mathbb{R}$ that FPA-rationalizes the data set.This utility function $u(x)$ will be differently defined on three parts of the domain $\mathbb{R}$ (labeled by w_1, w_2 and w_3). One part for medium to large values of x. A second part for small values of x and a final part for negative values of x. The reason for splitting up the domain is to make sure that $u(0) = 0$.

Given that G is Lipschitz continuous, we have that H is also Lipschitz continuous. Let λ be the Lipschitz constant on the cumulative distribution function H, that is, for all $x, y \in [0, \overline{v}]$,

$$|H(x) - H(y)| < \lambda |x - y|.$$

The cut-off for the domain of our utility function is given by the following technical lemma.

Lemma 2. *Assume that, for all $t = 1, \dots, T$, $v^t > 0$. Then there exists a number $0 \leq \underline{x} \leq \overline{v}$ such that, for all $t = 1, \dots, T$,*

$$\underline{x} \leq \frac{H(v^t - \underline{x})}{\lambda}.$$

Proof. Let $t \in \{1, \dots, T\}$. As $v^t > 0$ we have that $H(v^t) > 0$. Continuity of H implies that there exists ε^t, small enough, such that

$$H(v^t - \varepsilon^t) > 0.$$

Define

$$\underline{x} = \min_{t=1,\dots,T} \{\varepsilon_t, \frac{H(v^t - \varepsilon^t)}{\lambda}\}.$$

Then $\underline{x}$ is clearly positive and

$$\underline{x} \leq \min_{t=1,\dots,T} \frac{H(v^t - \varepsilon^t)}{\lambda} \leq \min_{t=1,\dots,T} \frac{H(v^t - \underline{x})}{\lambda}.$$

□

Let U^t solve the inequalities in (2). By rescaling all values of U^t, we can require that $U^t \leq 1$ for all t. Let $\underline{x}$ be as in the lemma above and assume without loss of generality that $\underline{x} \leq x^t$ for all $t = 1, \dots, T$ (since we can choose the ε_t arbitrarily small).

First part: $x \in [\underline{x}, \overline{v}]$. We set

$$w_1(x) = \min_{t=1,\dots,T} \left\{ \min \left\{ U^t \frac{H(v^t - x^t)}{H(v^t - x)}; 1 \right\} \right\},$$

where we define $\frac{a}{0} = +\infty$ for all $a > 0$. Given that $H(v^t - x^t) = H(b^t) > 0$, the right hand side is well defined. First, notice that $w_1(x)$ is the minimum of a finite number of continuous, non-decreasing functions, so it is continuous and non-decreasing.

Next, we need to show that the constructed utility function rationalizes the data. To do, we first argue that $w_1(x^t) = U^t$. Since the conditions of Theorem 1 are satisfied, we have that, for all observations s,

$$U^t \leq \min\left\{U^s \frac{H(v^s - x^s)}{H(v^s - x^t)}; 1\right\},$$

which shows that $w_1(x^t) = U^t$. Next we show that b^t is the optimal bid at observation t. If $H(v^t - x) > 0$, then by construction

$$w_1(x) \leq U^t \frac{H(v^t - x^t)}{H(v^t - x)},$$

and therefore

$$H(v^t - x)w_1(x) \leq H(v^t - x^t)U^t = H(v^t - x^t)w_1(x^t).$$

If $H(v^t - x) = 0$, then by default $H(v^t - x^t)U^t \geq H(v^t - x)w_1(x) = 0$ holds.

Second part: $x \in [0, \underline{x}[$. Let $\Delta = \frac{w_1(\underline{x})}{\underline{x}}$, and define

$$w_2(x) = \Delta x,$$

for every $x \in [0, \underline{x}[$. This construction guarantees that utility is nondecreasing, continuous and that $w_2(0) = 0$. Note that, since $U^t > 0$ and $H(v^t - x) \geq H(v^t - x^t) > 0$ for all t, we have that $w_2(\underline{x}) > 0$. As such, $\Delta > 0$ and $w_2(x) > 0$ for all $x \in]0, \underline{x}[$. Therefore, the utility function satisfies local non-satiation at zero.

We are left to show that $u(x)$ rationalizes the data. From above we have, for all $t = 1, \ldots, T$,

$$H(v^t - \underline{x})w_2(\underline{x}) \leq H(v^t - x^t)w_2(x^t).$$

Hence, it is sufficient to show that, for all $t = 1, \ldots, T$ and $x \in [0, \underline{x}[$,

$$H(v^t - x)w_2(x) \leq H(v^t - \underline{x})w_2(\underline{x}).$$

Due to Lipschitz continuity we have

$$\begin{aligned} H(v^t - x)w_2(x) &= \frac{w_2(\underline{x})}{\underline{x}} x H(v^t - x) \\ &\leq w_2(\underline{x}) \frac{x}{\underline{x}} \left(H(v^t - \underline{x}) + \lambda(\underline{x} - x)\right). \end{aligned}$$

This implies

$$\begin{aligned}
&w_2(\underline{x})\frac{x}{\underline{x}}\left(H(v^t-\underline{x})+\lambda(\underline{x}-x)\right) \le w_2(\underline{x})H(v^t-\underline{x})\\
\Leftrightarrow& w_2(\underline{x})\left(H(v^t-\underline{x})\left(\frac{x}{\underline{x}}-1\right)+\lambda\left(1-\frac{x}{\underline{x}}\right)x\right)\le 0\\
\Leftrightarrow&\left(1-\frac{x}{\underline{x}}\right)\left(\lambda x-H(v^t-\underline{x})\right)\le 0\\
\Leftrightarrow&\lambda x-H(v^t-\underline{x})\le 0\\
\Leftrightarrow& x\le\frac{H(v^t-\underline{x})}{\lambda}.
\end{aligned}$$

The latter holds since $x \le \underline{x} \le \frac{H(v^t-\underline{x})}{\lambda}$

Third part: $x \le 0$. The we can define $w_3(x) = x$, which guarantees that $H(v^t - x)w_3(x) \le 0$ and it is never optimal to set $x \le 0$.

B Robustness checks

In this appendix, we present two robustness checks of our empirical analysis in Section 6. Particularly, in our main analysis we have split the data per subject into five blocks of 10 rounds each. In what follows, we redo our statistical tests for blocks consisting of 8 and 12 rounds. This will show that our main conclusions are not overly reliant on our particular choice of block length.

Table 4 presents the results for the setting with 8 round blocks. Since we shorten the time period, our test uses less inequality constraints per subject, which makes the test less powerful. In turn, this results in larger differences between the pass rates for generic significance levels (first two columns) and power-adjusted significance levels (last two columns). Importantly, however, we do find again that (almost all) all pass rates are higher for the "real" subjects than for the "randomly generated" subjects. Also, we observe once more the increasing pass rates from earlier to later blocks, especially in the last two columns. The results for the first two columns are somewhat more noisy. In particular, pass rates for the real subjects are not significantly different from pass rates for the random subjects for the third, fourth and fifth blocks when using the generic significance levels. In addition, there is no clearly increasing trend in pass rates over time for these significance levels: 24 subjects pass our test of FPA-rationalizability in the second block, while only 12 subjects pass the test in the third block. Thus, when we decrease the length of the time blocks, the results for the generic significance levels are somewhat more noisy, and the power adjustment becomes more important.

Rounds		$\alpha \geq .05$	$\alpha \geq .10$	Power= .9	Power= .95
$1 \leq t \leq 8$	Pass Rate	8 (29%)	8 (29%)	8 (29%)	8 (29%)
	95% conf. interval	13% – 49%	13% – 49%	13% – 49%	13% – 49%
	Random subjects	46%	44%	10%	5%
$9 \leq t \leq 16$	Pass Rate	24 (86%)	24 (86%)	8 (29%)	8 (29%)
	95% conf. interval	67% – 96%	67% – 96%	13% – 49%	13% – 49%
	Random subjects	57%	55%	10%	5%
$17 \leq t \leq 24$	Pass Rate	12 (43%)	12 (43%)	8 (29%)	8 (29%)
	95% conf. interval	25% – 63%	25% – 63%	13% – 49%	13% – 49%
	Random subjects	40%	40%	10%	5%
$25 \leq t \leq 32$	Pass Rate	16 (57%)	16 (57%)	12 (43%)	8 (29%)
	95% conf. interval	37% – 76%	37% – 76%	25% – 63%	13% – 49%
	Random subjects	47%	44%	10%	5%
$33 \leq t \leq 40$	Pass Rate	20 (71%)	20 (71%)	12 (43%)	12 (43%)
	95% conf. interval	51% – 87%	51% – 87%	25% – 63%	25% – 63%
	Random subjects	56%	55%	10%	5%
$41 \leq t \leq 48$	Pass Rate	28 (100%)	28 (100%)	20 (43%)	12 (43%)
	95% conf. interval	88% – 100%	88% – 100%	51% – 87%	25% – 63%
	Random subjects	46%	44%	10%	5%

Table 4: Non-rejection rates for H_0. Time blocks of 8 periods

Table 5 presents the results for blocks consisting of 12 rounds. The pass rates are consistent with the results presented in Section 6 of the main text. In particular, the pass rates for the real subjects are higher than for the random subjects in all cases. In addition, there is an increasing trend in pass rates for real subjects from the beginning to the end of the experiment. Once more, our results confirm that block length impacts the noisiness of our estimates.

Rounds		$\alpha \geq .05$	$\alpha \geq .10$	Power= .9	Power= .95
$1 \leq t \leq 12$	Pass Rate	20 (71%)	20 (71%)	12 (43%)	12 (43%)
	95% conf. interval	51% – 87%	51% – 87%	25% – 63%	25% – 63%
	Random subjects	35%	34%	10%	5%
$13 \leq t \leq 24$	Pass Rate	20 (71%)	20 (71%)	16 (57%)	12 (43%)
	95% conf. interval	51% – 87%	51% – 87%	37% – 76%	25% – 63%
	Random subjects	44%	38%	10%	5%
$25 \leq t \leq 36$	Pass Rate	20 (71%)	20 (71%)	16 (57%)	12 (43%)
	95% conf. interval	51% – 87%	51% – 87%	47% – 76%	25% – 63%
	Random subjects	49%	48%	10%	5%
$37 \leq t \leq 48$	Pass Rate	24 (86%)	24 (86%)	20 (71%)	20 (71%)
	95% conf. interval	67% – 96%	67% – 96%	51% – 87%	51% – 87%
	Random subjects	41%	40%	10%	5%

Table 5: Non-rejection rates for H_0. Time blocks of 12 periods