Estimation of Games under No Regret: Structural Econometrics for AI*

Niccolò Lomys[†] Lorenzo Magnolfi[‡]

November 29, 2024

Abstract

We develop a method to recover primitives from data generated by artificial intelligence (AI) agents in strategic environments such as online marketplaces and auctions. Building on how leading online learning AIs are designed, we assume agents minimize their regret. Under asymptotic no regret, we show that time-average play converges to the set of Bayes coarse correlated equilibrium (BCCE) predictions. Our econometric procedure is based on BCCE restrictions and convergence rates of regret-minimizing AIs. We apply the method to pricing data in a digital marketplace for used smartphones. We estimate sellers' cost distributions and find lower markups than in centralized platforms.

Keywords: AI Decision-Making; Empirical Games; Regret Minimization; Bayes (Coarse) Correlated Equilibrium; Partial Identification.

JEL Classification: C1; C5; C7; D4; D8; L1; L8.

*We received helpful feedback from Ali Aouad, Martino Banchio, Panle Jia Barwick, Dirk Bergemann, Steve Berry, Ben Brooks, Emilio Calvano, Giacomo Calzolari, Áureo de Paula, Daniel Ershov, Mira Frick, Cristina Gualdani, Karam Kang, Ryota Iijima, Konrad Menzel, Francesca Molinari, Juan Ortner, Dan Quint, Jérôme Renault, Karl Schlag, Eran Shmaya, Alan Sorensen, Jörg Stoye, Chris Sullivan, Vasilis Syrgkanis, Éva Tardos, Emanuele Tarantino, and Kangning Wang. We thank Camilla Roncoroni for input on an earlier version of this manuscript. We are grateful to various seminar and conference audiences for their comments. We thank the NET Institute (www.NETinst.org) for financial support. Niccolò Lomys acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreements № 714147 and № 714693). Part of this work was done while Niccolò Lomys was visiting the Simons Institute for the Theory of Computing at UC Berkeley (program on Data-Driven Decision Processes).

†CSEF and Università degli Studi di Napoli Federico II; niccolomys@gmail.com.

‡Department of Economics, University of Wisconsin-Madison; magnolfi@wisc.edu.

1 Introduction

Firms increasingly delegate strategic decisions to artificial intelligence (AI) or AI-aided agents (see, e.g., Chen, Mislove, and Wilson, 2016). This shift is particularly evident in digital marketplaces, where sellers use algorithms to set prices on platforms like Amazon Marketplace and Walmart Marketplace. Similarly, in online advertising, automated bidding systems compete for ad placements on platforms like Bing Ads and Google DoubleClick. The rise of algorithmic decision-making generates crucial questions about competition policy and market design (Milgrom and Tadelis, 2018; Banchio and Skrzypacz, 2022), many of which require estimating economic primitives to evaluate policy counterfactuals.

Modeling data generated by strategic interactions among algorithms calls for new empirical tools. AI agents adopt sophisticated algorithms to deal with the exploration-exploitation trade-off typical of *online learning* problems. In online learning, agents with limited knowledge of the environment choose actions in each period and receive payoffs that can depend arbitrarily on their opponents' choices and a state of nature. By taking new actions, agents learn their associated payoffs (exploration) but at the cost of potentially suboptimal choices given current information (exploitation). *Regret minimization* has emerged in computer science and algorithmic game theory as the leading paradigm to design online learning algorithms and evaluate their performance (e.g., Cesa-Bianchi and Lugosi, 2006; Nisan, Roughgarden, Tardos, and Vazirani, 2007; Roughgarden, 2016). Regret is the difference between the time-average payoff an agent could have obtained by repeatedly choosing the best fixed action in hindsight in all past periods and the time-average payoff the agent actually obtained. State-of-the-art learning algorithms are designed around the regret-minimization objective (Slivkins, 2019; Lattimore and Szepesvári, 2020; Hazan, 2022): in the long run, agents should do at least as well as with the best fixed action in hindsight.[1]

In this paper, we develop a method to estimate economic primitives in environments where AI supports or replaces human decisions, building on the theoretical properties of regret-minimizing algorithms. We model agents' interaction in an incomplete information environment as an online learning problem; consistent with the theory and practice of online learning, we impose asymptotic no regret (ANR) as a minimal optimality condition on the behavior observed in the data. Hence, our model of AI decision-making departs from standard structural econometric methods, which rely on equilibrium models where agents must have prior knowledge or beliefs about their opponents' identities, incentives, and information.

ANR is an optimality property of observed behavior but does not specify a decision rule

[1]That is, agents should do at least as well as with the best fixed action had they known the empirical distribution of payoffs associated with each action in advance.

nor requires agents to compute regrets each period. Leading AI algorithms guarantee ANR under minimal informational requirements—i.e., knowing only their action set and estimates of realized payoffs (e.g., Shalev-Shwartz, 2012; Bubeck and Cesa-Bianchi, 2012). Although we do not exclude that agents may know or observe more or that they might coordinate on specific algorithms, such additional assumptions are unnecessary. Hence, our econometric approach is in the spirit of incomplete models (Tamer, 2003; Haile and Tamer, 2003).

Under the ANR assumption, we develop an empirical strategy to estimate the structural parameters of the underlying stage game. To characterize the empirical content of ANR, we establish an auxiliary result on the convergence of regret-minimizing dynamics: ANR holds *if and only if* the time average of play (i.e., the empirical distribution of actions over time) converges to the set of Bayes coarse correlated equilibrium (BCCE) predictions of the stage game.

A BCCE is a joint probability distribution over actions, types, and states satisfying two restrictions: consistency and coarse obedience. Consistency requires that integrating out actions (i.e., the endogenous variables) from the BCCE distribution recovers the true distribution of types and states (i.e., the exogenous variables). Coarse obedience requires that when an action profile, type profile, and state are drawn from the BCCE distribution, each agent, who is informed only of their realized type from the draw, prefers to commit to their part of the realized action profile rather than to any other action. Our BCCE notion is the coarse analog of the Bayes correlated equilibrium (BCE) notion of Bergemann and Morris (2016). The novel result on convergence under ANR provides dynamic foundations for the BCCE notion, thus generalizing to incomplete information environments earlier findings for games with complete information (e.g., Foster and Vohra, 1997; Hart and Mas-Colell, 2000).

The intuition for the convergence result under ANR is that if the regret of each type of each agent vanishes in the long run, the time average of play must eventually satisfy the incentive constraints corresponding to coarse obedience, and vice versa. Coarse obedience, requiring only that agents' actions be optimal conditional on their types, reflects that regret-minimizing agents only consider "coarse" deviations to fixed alternative actions. Hence, the convergence of the empirical distribution of actions to the set of BCCE predictions provides us with static equilibrium restrictions (namely, consistency and coarse obedience) that we can use to identify and estimate games played by regret-minimizing AIs.

Identification in our setting differs from standard approaches where data are i.i.d. draws from a single limiting distribution. Under ANR, neither period-by-period actions nor their empirical distribution over time necessarily converge to a specific limiting distribution. Instead, the observed sequence of empirical distributions eventually enters any neighborhood of the set of BCCE predictions; thereafter, the sequence may wander forever within the neighborhood in arbitrary ways. For this reason, we define the identified set as those parameters that rationalize the observed empirical distribution of actions as generated by regret-minimizing behavior. The set is characterized by moment inequalities derived from the BCCE constraints. Because of the if-and-only-if nature of our convergence result, these inequalities provide sharp restrictions for identification under ANR and allow us to develop a tractable estimation procedure.

We further leverage the algorithmic nature of our data-generating process for inference. We construct uniformly valid confidence regions using theoretical bounds on the convergence rates of regret-minimizing algorithms to their objective. These bounds can be obtained under conservative assumptions about the environment, i.e., only requiring algorithms to operate in an adversarial setting with minimal feedback about their performance. Alternatively, researchers can obtain tighter confidence regions by imposing additional assumptions, e.g., maintaining that agents receive richer feedback about their payoffs or that they adopt regret-minimizing algorithms in some specific sub-class. This flexibility allows researchers to tailor inference to their empirical context, trading off the robustness of conservative assumptions against the increased precision available under stronger ones. We illustrate our estimation approach and these trade-offs through Monte Carlo simulations of a repeated pricing game.

In an empirical application, we use our method to study competition between sellers on Swappa, a decentralized marketplace for used smartphones. Unlike standard empirical pricing models, we do not posit that sellers know or correctly conjecture their competitors' strategies; instead, we assume that sellers adopt ANR pricing algorithms. In this setting, ANR requires that sellers could not have obtained systematically higher profits by committing to any fixed price for any of their marginal costs. Under ANR, our method allows us to use BCCE restrictions to recover the distribution of marginal costs for the platform's largest sellers, who make frequent pricing decisions and are most likely to employ algorithmic pricing tools.

We find that these sellers' marginal costs are \$56-\$90 below prices on centralized platforms. Using these estimates, we quantify markups and compare them to those charged by Gazelle, a centralized buyback and resale platform. Our analysis reveals that margins for Swappa sellers are modest, with average markups of \$21-\$51 per device—substantially lower than those of centralized resellers, especially for newer devices. These findings provide new evidence on the profitability of different business models in e-commerce and speak to the broader debate on the relative competitiveness of first-party versus third-party sellers.

While our application recovers costs and markups from pricing data, our method extends naturally to other settings where strategic agents employ learning algorithms. The approach can recover any primitive that affects payoffs—whether marginal costs, utility parameters, or other structural features. As AI tools become increasingly prevalent in market interactions, this flexibility makes our framework portable and useful across empirical environments.

Related Literature. We contribute to a recent literature that estimates empirical models of learning agents (see Aguirregabiria and Jeon, 2020, for a survey). Existing work adopts model-based and belief-based learning approaches (e.g., Doraszelski, Lewis, and Pakes, 2018; Aguirregabiria and Magesan, 2020). In contrast, conforming to the practice of algorithmic decision-making in strategic environments, our AI agents need not model or form beliefs about other agents' behavior and information. By relying only on a principle (ANR) that guides the design of online learning algorithms, our approach leads naturally to an incomplete model and results in partial identification (see Molinari, 2020, for a survey).

Although we use related equilibrium restrictions, we depart from the literature that uses BCE for inference (Syrgkanis, Tamer, and Ziani, 2021; Magnolfi and Roncoroni, 2023; Gualdani and Sinha, 2024) by using BCCE restrictions to capture long-run outcomes of regret-minimizing AIs, without explicit focus on the robustness to informational assumptions. This yields a distinct econometric approach: we establish inferential properties using convergence rates of regret-minimizing algorithms rather than assuming that data are i.i.d. samples from an equilibrium distribution.

Our paper relates to an emerging literature that studies the implications of algorithmic decision-making on competition. Existing work used simulation (e.g., Calvano, Calzolari, Denicolò, and Pastorello, 2020; Asker, Fershtman, and Pakes, 2024), theoretical (e.g., Salcedo, 2015; Hansen, Misra, and Pai, 2021; Aouad and den Boer, 2021; Brown and MacKay, 2023; Lamba and Zhuk, 2024; Banchio and Mantegazza, 2024), and empirical approaches (e.g., Musolff, 2024; Assad, Clark, Ershov, and Xu, 2024) to shed light on the potential for algorithms to sustain supra-competitive prices. We complement these theoretical and simulation studies by proposing a structural econometrics approach.

The existing literature mainly focuses on AIs based on reinforcement learning (RL), which may sustain supra-competitive prices. Instead, regret-minimizing AIs pursue approximate maximization of short-run profits, adapting to market conditions. As a result, no-regret properties have been proposed to define non-collusive outcomes of algorithmic competition (Chassang and Ortner, 2023), and deviations from no regret may form a basis for detecting collusion in AI pricing data (Hartline, Long, and Zhang, 2024). The theoretical distinction between regret minimization and collusion suggests an empirical distinction between two environments where agents may deploy AIs. In concentrated environments where few agents compete, algorithmic collusion is a first-order concern. Our method, instead, mainly applies in decentralized environments, where no regret is a sensible benchmark for AIs, and even RL fails to reach collusive outcomes (see, e.g., the local learning case in Abada and Lambin, 2023).

Our work is related to the path-breaking study of Nekipelov, Syrgkanis, and Tardos (2015) and other recent work on econometric approaches to sponsored search auctions based on regret-minimizing bidding algorithms (Noti and Syrgkanis, 2021; Chen, Nabi, and Siniscalchi, 2023). We share with these authors the motivation for using regret minimization as a basis for empirical work, but our approach differs in two important ways. First, while Nekipelov et al. (2015) rely on the structure of second-price auctions, we use theoretical convergence results to characterize the empirical content of ANR through the static BCCE notion. Thus, our method applies to any underlying stage game and accommodates a broader class of information and stochastic environments. Second, while they focus on recovering the smallest regret error compatible with the data, we focus on estimating structural parameters. Since agents may still be learning in the data we observe, we leverage discrepancies from perfect regret minimization and the convergence rates of regret-minimizing AIs to conduct inference in finite samples.

Finally, we build on a large and growing literature at the intersection of economics and computer science on learning and regret minimization. Regret minimization is a leading approach in online learning and multi-armed bandit problems (see, e.g., Foster and Vohra, 1999; Slivkins, 2019; Lattimore and Szepesvári, 2020, and references therein) and a central idea in multiagent learning (see, e.g., Nisan et al., 2007; Roughgarden, 2016; Hart and Mas-Colell, 2013, and references therein). We contribute to this work by providing novel results on the convergence properties of regret-minimizing dynamics in games with incomplete information.

Road Map. In Section 2, we present the empirical model and the econometric problem we consider. In Section 3, we formalize the notion of asymptotic no regret and study convergence to the set of BCCEs. In Section 4, we develop our econometric approach. In Section 5, we present our empirical application. In Section 6, we conclude. All proofs are in Appendix A.

2 Empirical Model

In this section, we present the strategic interactions we consider and define asymptotic no regret. Then, we introduce the econometric problem and the data-generating process.

2.1 Strategic Interaction

Dynamic Environment. We model the strategic interaction of artificial intelligence (AI) or AI-aided agents (Immorlica, Lucier, and Slivkins, 2024) as an *online learning problem.* Online learning is the standard framework in computer science and algorithmic game theory to model sequential decision-making in changing environments under poor prior knowledge (see, e.g., Cesa-Bianchi and Lugosi, 2006; Nisan et al., 2007; Roughgarden, 2016). In online learning, each period, an agent probabilistically chooses an action, Nature makes its move, and the agent receives a payoff as a function of his action and Nature's move. The sequence of payoffs the agent receives over time can be arbitrary, need not be stationary, and might be adaptive to the agent's past actions. Agents receive information about past payoffs and must make decisions without strong knowledge or assumptions about the future payoffs' distribution. In a strategic interaction, each period's payoffs are a function of the agent's actions, the actions of his opponents, and stochastic payoff-relevant features of the environment; together, the latter two elements act as Nature's move.

Formally, we consider finitely many agents, $i \in \mathcal{I} \coloneqq \{1, \dots, I\}$, interacting over discrete-time periods, $n \in \mathbb{N} \coloneqq \{1, 2, \dots\}$. In each period n:

(i) A payoff state θ_n is drawn from some state distribution $\psi_n \in \Delta(\Theta)$, where Θ is the finite set of payoff states.

(ii) A profile of signals $t_n \coloneqq (t_{1,n}, \dots, t_{I,n}) \in T \coloneqq T_1 \times \cdots \times T_I$ is drawn from some signal distribution $\pi(\cdot \mid \theta_n) \in \Delta_{++}(T)$, where T_i is the finite set of signals (types) of agent i. Each agent i privately observes his signal $t_{i,n}$.

(iii) Each agent i selects an action $a_{i,n} \in A_i$, where A_i is the finite set of actions of agent i.

(iv) Each agent i observes his realized payoff $u_i(a_n, \theta_n)$, where $a_n \coloneqq (a_{1,n}, \dots, a_{I,n}) \in A \coloneqq A_1 \times \cdots \times A_I$ is the action profile played in period n.

We refer to the resulting dynamic strategic interaction as the *dynamic environment* and denote it by $\mathcal{G}^\infty$. We specify agents' behavior in $\mathcal{G}^\infty$ in Section 2.2.

A history of $\mathcal{G}^\infty$ at the end of period N is a sequence $(a_n, t_n, \theta_n)_{n=1}^N$. Let $H^N \coloneqq (A \times T \times \Theta)^N$ be the set of histories at the end of period N, $H \coloneqq \cup_{N \geq 1} H^N$ the set of finite histories, and $H^\infty \coloneqq (A \times T \times \Theta)^\infty$ the set of infinite histories. We refer to $(a_n, t_n, \theta_n)_{n=1}^\infty \in H^\infty$ as a *sequence of actions, signals, and states from* $\mathcal{G}^\infty$.

We denote by θ, a_i, a, t_i, and t typical elements of (sub)sets Θ, A_i, A, T_i, and T. We write $\pi(t \mid \theta)$ for the probability of signal profile t when the payoff state is θ. We denote by a_{-i} (resp., t_{-i}) a typical profile of actions (resp., signals) for agents other than i.

Our general specification of a dynamic environment makes no assumptions about the evolution of payoff states. In the main body of the paper, to help intuition, we focus on an environment directly relevant to many empirical applications, including ours: payoff states θ_n are i.i.d. across periods, drawn from some state distribution $\psi \in \Delta_{++}(\Theta)$. Although all main features of the method are well-illustrated in this setting, our approach applies more broadly. We discuss more general dynamic environments in Appendix B.

We assume that each agent knows his set of possible actions and, in each period, observes his signal and realized payoff. However, we make no further knowledge or monitoring assumptions. Agents may not know their opponents' identities and payoffs, how payoff states evolve, and how states map into signals. Thus, agents may not have a (common) prior, may not process the information that signals convey, and may not observe the state and the action profile at the end of each period. Importantly, we do not exclude that agents know or observe more; such additional requirements, however, are unnecessary for our results.

Stage Game. To the dynamic environment $\mathcal{G}^\infty$, there corresponds an *(incomplete information) static game* $\mathcal{G} \coloneqq (\mathcal{I}, G, S)$, where G is the basic game and S is the information structure. The *basic game* G consists of: (i) the set of payoff states Θ; (ii) for each agent i, the set of actions, A_i, and the payoff function, $u_i\colon A \times \Theta \to \mathbb{R}$; (iii) the state distribution $\psi \in \Delta_{++}(\Theta)$. Thus, $G \coloneqq \big(\Theta, (A_i, u_i)_{i=1}^I, \psi\big)$. The *information structure* S consists of: (i) for each agent i, the set of signals (types) T_i; (ii) the signal distribution $\pi\colon \Theta \to \Delta_{++}(T)$. Thus, $S \coloneqq \big((T_i)_{i=1}^I, \pi\big)$. If $|\Theta| = 1$, then $\mathcal{G}$ is a game with complete information. Hereafter, we refer to $\mathcal{G}$ as the *stage game*.

Structural Parameters. The stage game associated with the dynamic environment $\mathcal{G}^\infty$ belongs to a parametric class $\{\mathcal{G}(\lambda)\}_{\lambda\in\Lambda}$, indexed by the *structural parameters* $\lambda \in \Lambda \neq \emptyset$.

Discussion. Our setting departs from typical models of empirical games to capture AI decision-making in decentralized online environments such as digital marketplaces and sponsored search auctions. In these complex environments, agents typically make high-frequency decisions in an evolving landscape. Hence, agents find it hard to know all the relevant features of their interaction, such as having prior knowledge or forming (correct) beliefs about their opponents' identities, incentives, information, and behavior. Therefore, although we use standard game-theoretic language to present our model, game-theoretic formalisms, like the basic game and the information structure, do not have the usual interpretation in our setting: they do not represent agents' actual knowledge and beliefs about the environment. Nonetheless, game-theoretic notions and restrictions on the observables will be key ingredients of our econometric procedure.

2.1.1 Illustration: A Two-Seller Pricing Game

Consider pricing by firms in an e-commerce platform, the environment we study in our empirical application. In this and our following illustrations, we use a stylized version of the environment to introduce the main elements of our method in a simple setting.

There are two sellers, $i \in \mathcal{I} = \{1, 2\}$, each with one unit of a differentiated good. Sellers rely on AI pricing algorithms and interact over discrete-time periods $n \in \mathbb{N}$. In each period n:

(i) A profile of marginal costs $\theta_n = (t_{1,n}, t_{2,n})$, i.i.d. over time, is drawn from some probability distribution $\psi(\cdot; \lambda) \in \Delta_{++}(\Theta)$, where $\Theta = T = T_1 \times T_2$ and $1 \leq |\Theta| < \infty$.

(ii) Each seller i privately observes his marginal cost $t_{i,n} \in T_i$. Hence, the information structure S has $\pi(t \mid \theta; \lambda) = 1$ if and only if $t = \theta$.

(iii) Each seller i sets price $p_{i,n} \in A_i = \{p_\ell, p_h\}$, where $0 < p_\ell < p_h$.

(iv) Each seller i observes his profit $u_i((p_i, p_{-i}), \theta; \lambda) = s_i(p_{i,n}, p_{-i,n}; \lambda)(p_{i,n} - t_{i,n})$, where $s_i(p_{i,n}, p_{-i,n}; \lambda)$ is the probability of sale for i's good when prices are $(p_{i,n}, p_{-i,n})$.

Each seller knows the prices he can set and observes his marginal cost and realized profit in each period. We do not make further knowledge or monitoring assumptions.

2.2 Asymptotic No Regret

In online learning, agents face the exploration-exploitation trade-off typical of bandit problems: taking new actions allows agents to learn the payoffs associated with them (exploration) but at the cost of sub-optimal behavior based on the information accumulated so far (exploitation). Regret minimization, i.e., outperforming a benchmark policy in hindsight, is the standard objective to evaluate performance in online learning problems. In particular, minimizing regret with respect to the best fixed action in hindsight captures the best agents can consistently achieve in environments about which prior knowledge may be hard to obtain. As such, this worst-case objective is the principle guiding the design of online learning algorithms (see, e.g., Shalev-Shwartz, 2012; Bubeck and Cesa-Bianchi, 2012).

To reflect these properties of online learning (or bandit) algorithms, we assume that agents' behavior in $\mathcal{G}^\infty$ satisfies the minimal optimality condition of *asymptotic no regret*, which we now formally define. For all $i \in \mathcal{I}$ and $t_i \in T_i$, let $U_N(i, t_i)$ be the average *factual payoff* that agent i with signal t_i has obtained up to period N:

$$U_N(i, t_i) \coloneqq \frac{1}{N} \sum_{n=1}^{N} u_i\big((a_{i,n}, a_{-i,n}), \theta_n; \lambda\big) \mathbb{1}_{\{t_i\}}(t_{i,n}),$$

where $\mathbb{1}_{\{.\}}(\cdot)$ is the indicator function. Moreover, let $V_N(i, t_i)$ be the average *counterfactual payoff* that agent i with signal t_i would have obtained had he played the best fixed action in hindsight up to period N:

$$V_N(i, t_i) \coloneqq \max_{a_i \in A_i} \left\{ \frac{1}{N} \sum_{n=1}^{N} u_i\big((a_i, a_{-i,n}), \theta_n; \lambda\big) \mathbb{1}_{\{t_i\}}(t_{i,n}) \right\}$$

Regrets are defined as differences between these counterfactual and factual payoffs.

Definition 1 (Regret)**.** *For all $i \in \mathcal{I}$ and $t_i \in T_i$, the regret of agent i with signal t_i before play in period $N + 1$, denoted by $R_N(i, t_i)$, is defined as*

$$R_N(i, t_i) \coloneqq \max\{V_N(i, t_i) - U_N(i, t_i), 0\}.$$

$R_N(i, t_i)$ is the difference between (i) what agent i with signal t_i could have gotten had he known in advance the empirical distribution of payoffs associated with each action up to period N and chosen the best fixed action accordingly, and (ii) what he actually got.[2]

Asymptotic no regret requires the time-average regret experienced by each type of each agent for not having played the best fixed action in hindsight to vanish in the long run.

Definition 2 (Asymptotic No Regret). *A sequence $(a_n, t_n, \theta_n)_{n=1}^{\infty}$ from $\mathcal{G}^{\infty}$ has asymptotic no regret (ANR) if, for all $i \in \mathcal{I}$ and $t_i \in T_i$, we have*

$$\limsup_{N\to\infty} R_N(i, t_i) \leq 0 \qquad \textit{almost surely.}$$

Intuitively, ANR requires the no ex-post regret property of Nash equilibrium to hold over sequences $(a_n, t_n, \theta_n)_{n=1}^{\infty}$ with respect to the best fixed action in hindsight. In contrast to Nash equilibrium, however, ANR does not require agents' actions to be stable and independent. Moreover, ANR is weaker than equilibrium play: should agents reach or play any Bayes Nash or Bayes (Coarse) Correlated equilibrium of $\mathcal{G}$, the resulting sequence $(a_n, t_n, \theta_n)_{n=1}^{\infty}$ from $\mathcal{G}^{\infty}$ has ANR (Section 3 formalizes this idea).

Probability Space. Regret-minimizing algorithms—hereafter, *ANR algorithms*—typically involve randomization. Thus, there are three sources of randomness in the model: (i) the process of payoff states $(\theta_n)_{n=1}^{\infty}$; (ii) the sequence of signal distributions $(\pi(\cdot \mid \theta_n; \lambda))_{n=1}^{\infty}$; (iii) the randomization induced by ANR algorithms. These sources induce a probability measure $\mathbb{P}$ on the set of finite histories H. By the Kolmogorov extension theorem, $\mathbb{P}$ uniquely extends to H^{∞}. Hereafter, all probabilistic statements refer to the probability measure $\mathbb{P}$.

2.2.1 ANR Algorithms

ANR algorithms are data-driven procedures that learn from past experiences and adapt behavior to the environment while making decisions. Since the seminal work of Blackwell (1956a,b) and Hannan (1957), many ANR algorithms have been developed. Some popular ANR algorithms are: the Regret Matching algorithm (Hart and Mas-Colell, 2000); the Multiplicative Weights Update algorithm (Arora, Hazan, and Kale, 2012) and its precursors, such as the Weighted Majority algorithm (Littlestone and Warmuth, 1994) or the Hedge algorithm (Freund and Schapire, 1997); the Exponential-Weight algorithm for Exploration and Exploitation (Auer, Cesa-Bianchi, Freund, and Schapire, 2002); the Follow the Leader algorithm (Kalai and Vempala, 2005), and the Mirror Descent algorithm (Nemirovsky and Yudin, 1983).

Many ANR algorithms are computationally efficient. They do not need offline training and are designed to keep learning in constantly changing or adversarial environments. ANR algorithms require almost no prior knowledge, consistent with the minimal requirements we

[2]We discuss alternative regret notions in Appendix C.

impose in Section 2.1 on what agents know or observe about their interaction. ANR algorithms can work even if counterfactual payoffs and regrets are not observable, as these can be estimated via experimentation. Moreover, ANR algorithms are model- and belief-free: they do not require agents to form a model or beliefs about the environment, their opponents' objectives and play, and to reply optimally to such conjectures.

Different ANR algorithms are optimized to perform under different environments, constraints, or objectives. Following the literature on online learning and multi-armed bandit algorithms, we can broadly classify ANR algorithms depending on the following features (for further details, we refer to Slivkins, 2019; Lattimore and Szepesvári, 2020; Hazan, 2022).

1. Environment. What generates payoffs? The literature considers two extremes. In the *stochastic stationary bandit* environment, the payoff from each action is drawn independently from a fixed distribution that depends on the action but not on the period, the previous actions, and the previous payoffs. In the *adversarial bandit* environment, payoffs can be arbitrary, as if an adversary chooses the ANR algorithm's payoffs trying to fool the algorithm (without knowing the algorithm's action for the period). There are all shades of gray between these two extremes. The "less adversarial" the environment is, the easier it is to satisfy the desired performance benchmark for the ANR algorithm.

2. Informational Feedback. What feedback is available to the ANR algorithm after each period? The literature distinguishes three types of feedback. Under *full feedback*, the algorithm perfectly observes the payoffs for the selected action and all actions it could have selected instead. Under *bandit feedback*, the algorithm perfectly observes the payoff for the selected action but no other feedback. Under *partial feedback*, the algorithm observes a signal that may not be perfectly informative of the payoff of the selected action; such signal, however, may also provide some information about unselected actions, but not necessarily their exact payoffs.[3] In between, there are many other possible feedback structures. If richer feedback is available, it helps improve the performance of the ANR algorithm.

3. Objective. How does the ANR algorithm pursue its goal? For a given environment and informational feedback, many ANR algorithms minimize regret while maximizing the convergence rate of the expected regret to 0. The literature has characterized lower bounds for such a rate and various algorithms that achieve those bounds. In specific applications, other objectives may be desirable. For instance, some ANR algorithms minimize regret while optimizing properties of the distribution of realized regrets, e.g., mean-variance trade-off considerations.

For our theoretical and identification results, we only impose that agents' behavior sat-

[3] To ease exposition, we presented our model assuming that each agent observes at least his realized payoff at the end of each period. All our results, however, hold with no modification under partial feedback.

isfies ANR. Thus, our approach is *algorithm-independent*: we assume neither that agents adopt a specific ANR algorithm nor that they coordinate on the same one. However, we will exploit the properties of broad classes of ANR algorithms to establish inferential properties of our confidence regions.

The flexibility and appealing theoretical properties of ANR algorithms have led to their widespread adoption. Algorithms with regret minimization at their core are applied to bidding in online advertising auctions, recommender systems, pricing, and revenue management (see Slivkins, 2019, for more examples and references).

2.2.2 Illustration: A Two-Seller Pricing Game

Consider a parametric and numerical specification of the two-seller pricing game in Section 2.1.1. Sellers set prices $p_{i,n} \in \{p_\ell, p_h\} = \{4, 8\}$. Marginal costs are i.i.d. across sellers. Each seller's marginal costs are drawn from a discretized Normal distribution with parameters $\mu = 3$ and $\sigma = 1$, and support over 20 equally spaced points between $\underline{t} = 0$ and $\bar{t} = 6$. Let $s_i(p_{i,n}, p_{-i,n}; \lambda) := \exp(\eta p_{i,n})/[1 + \exp(\eta p_{i,n}) + \exp(\eta p_{-i,n})]$ be the probability of sale for seller i's good when prices are $(p_{i,n}, p_{-i,n})$. Hence, $s_1 = s_2 = s$. The parameter $\eta = -1/3$ captures consumers' price sensitivity so that i's selling probability decreases in $p_{i,n}$ and increases in $p_{-i,n}$. With this specification, the pricing game is symmetric.

We simulate sellers setting prices using *regret matching* (Hart and Mas-Colell, 2000). According to regret matching, the larger the regret for not having always set a given price in the past, the larger the probability of setting that price in the current period. Formally, let

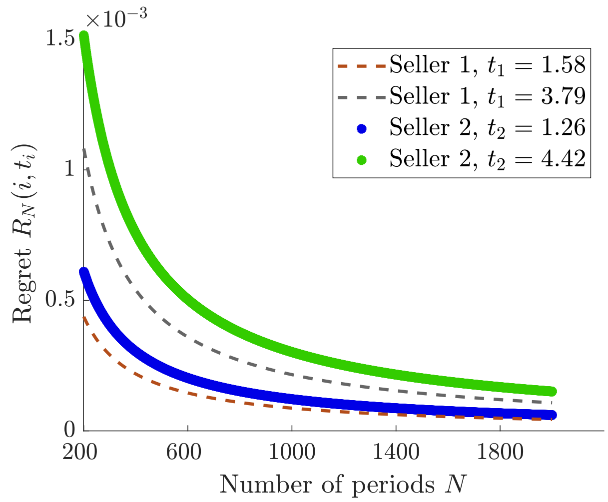
$$R_N(i, t_i, p) := \frac{1}{N} \sum_{n=1}^{N} \left[s(p, p_{-i,n}; \lambda)(p_i - t_{i,n}) - s(p_{i,n}, p_{-i,n}; \lambda)(p_{i,n} - t_{i,n}) \right] \mathbb{1}_{\{t_i\}}(t_{i,n})$$

be the regret of seller i with marginal cost t_i for not having always set price p up to period N. According to regret matching, the probability of setting price p in period $N+1$ by seller i with signal t_i, denoted by $\gamma_{N+1}(i, t_i, p)$, is proportional to the vector of his regrets, i.e., $\gamma_{N+1}(i, t_i, p) = R_N(i, t_i, p)/[R_N(i, t_i, 4) + R_N(i, t_i, 8)]$. Play is arbitrary when all regrets are 0.

Since a learning opponent's behavior is not guaranteed to follow some fixed stationary distribution, sellers face an adversarial bandit environment. Moreover, for simplicity, we implement regret matching by assuming counterfactual profits and regrets are observable, which requires sellers to know the demand function and observe their opponent's price each period. The full-feedback assumption, however, is easy to relax. Hart and Mas-Colell (2001) provide a modification of regret matching, dubbed *proxy-regret matching*, that works under bandit feedback, i.e., even if each seller is informed only of his realized profit in each period.

If both sellers set prices according to regret matching, the resulting sequence of prices

Figure 1: Regrets under Regret Matching.



We simulate price setting under regret matching for 10,000 periods. For two different marginal costs of each seller, we plot the regrets $R_N(i,t_i)$ as a function of the number of periods of play. We plot regrets starting in period 200 so that each seller has already played at least a few times for all marginal costs.

and marginal costs has ANR, as illustrated by Figure 1. For 10,000 iterations of regret matching, we represent in the figure how regrets $R_N(i,t_i)$ evolve for two different marginal costs of each seller as a function of the number of periods. For a given marginal cost, say t_i, how fast regrets converge to 0 depends on how often seller i sets prices for that specific cost (i.e., how frequently marginal cost t_i realizes) and the magnitude of the difference between payoffs corresponding to different profiles of actions.

2.3 Data-Generating Process

The true structural parameters λ_0 are unknown to a researcher who wants to learn about them. The assumption below summarizes the data-generating process and the observables.

Assumption 1. *Consider a dynamic environment $\mathcal{G}^\infty$ with stage game $\mathcal{G}(\lambda_0)$:*

1. *For some positive integer N, the researcher observes an empirical distribution of actions $q_N \in \Delta(A)$ from $\mathcal{G}^\infty$, where*

$$q_N(a) := \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}_{\{a\}}(a_n). \tag{1}$$

2. *The sequence of actions, signals, and states $(a_n, t_n, \theta_n)_{n=1}^\infty$ from $\mathcal{G}^\infty$ has ANR.*

The researcher may have access to more information than in Assumption 1. For instance, she may observe data on covariates or the full sequence $(q_n)_{n=1}^N$. We discuss in Appendix D how these additional observables help increase the informativeness of our econometric procedure. Furthermore, we note that the first period used to compute q_N need not coincide with the first period in $\mathcal{G}^\infty$ because the usual initial conditions problem in dynamic empirical models does not arise in our setting. Indeed, if a sequence $(a_n, t_n, \theta_n)_{n=1}^\infty$ from $\mathcal{G}^\infty$ has ANR, so does any other sequence $(a_n, t_n, \theta_n)_{n=k}^\infty$ from $\mathcal{G}^\infty$, for all positive integers k.

The empirical model described by Assumption 1 is *incomplete* in the sense of Tamer (2003) and Haile and Tamer (2003). Although we assume agents learn and adapt to the environment sufficiently well for ANR to hold, we do not specify other model elements. First, we do not make specific assumptions about the stochastic process generating agents' payoffs and what agents know or observe—they may know or observe more than the bare minimum for ANR to hold. Second, we neither assume that agents adopt a specific ANR algorithm nor that they coordinate on the same one.

In the following sections, we develop an econometric procedure to set-identify λ_0 under Assumption 1. To characterize the empirical content of ANR, we start by establishing a novel game-theoretic result on the convergence of the empirical distribution of actions q_N to the set of Bayes coarse correlated equilibrium (BCCE) predictions of the stage game $\mathcal{G}(\lambda_0)$. This auxiliary result will allow us to obtain sharp restrictions for identification under ANR and develop a tractable estimation procedure.

2.3.1 Illustration: A Two-Seller Pricing Game

For the pricing game in Sections 2.1.1 and 2.2.2, suppose the researcher can estimate "offline" the probability of sale $s(\cdot,\cdot)$ and, for some positive integer N, observes an empirical distribution of prices $q_N \in \Delta\big(\{4,8\}^2\big)$. Assume the sequence of prices and marginal costs satisfies ANR. If the researcher wants to recover the distribution of marginal costs, we have $\lambda = \psi$.

3 Convergence under Asymptotic No Regret

In this section, we establish a game-theoretical result connecting ANR to Bayes coarse correlated equilibrium. The result is instrumental in developing our econometric procedure.

3.1 Bayes Coarse Correlated Equilibrium

A Bayes coarse correlated equilibrium of the stage game $\mathcal{G}(\lambda)$ is a probability distribution over actions, signals, and states $\nu \in \Delta(A \times T \times \Theta)$ satisfying certain restrictions.

Definition 3 (Bayes Coarse Correlated Equilibrium)**.** *The probability distribution $\nu \in \Delta(A \times T \times \Theta)$ is a Bayes coarse correlated equilibrium (BCCE) of $\mathcal{G}(\lambda)$ if:*

1. ν is consistent for $\mathcal{G}(\lambda)$; that is, for all $t \in T$ and $\theta \in \Theta$, we have

$$\sum_a \nu(a, t, \theta) = \pi(t \mid \theta; \lambda)\psi(\theta; \lambda).$$

2. ν is coarsely obedient for $\mathcal{G}(\lambda)$; that is, for all $i \in \mathcal{I}$ and $t_i \in T_i$, we have

$$\sum_{a, t_{-i}, \theta} \left[u_i\big((a_i', a_{-i}), \theta; \lambda\big) - u_i(a, \theta; \lambda)\right]\nu(a, (t_i, t_{-i}), \theta) \leq 0 \quad \text{for all } a_i' \in A_i.$$

We denote by $E(\lambda)$ the set of BCCEs of $\mathcal{G}(\lambda)$.

Consistency is a feasibility constraint requiring the marginal of distribution ν on $T \times \Theta$ to be consistent with the description of game $\mathcal{G}(\lambda)$. *Coarse obedience* is an incentive constraint best understood with the mediator metaphor. Suppose a mediator draws an action profile, a signal profile, and a state from distribution ν. Agents know ν. The mediator informs each agent i about his realized signal t_i, but not about his realized action a_i, from ν. Next, the mediator gives each agent i a choice between (a) committing to whatever joint action profile (a_i, a_{-i}) has realized from ν, and (b) committing to any fixed action a_i'. The distribution ν is coarsely obedient if each agent weakly prefers (a) to (b), given that the other agents are committed to playing their part in the realized action profile.

The set $E(\lambda)$ is convex. Our BCCE notion is the coarse analog of the Bayes correlated equilibrium (BCE) notion of Bergemann and Morris (2016) and an incomplete information version of coarse correlated equilibrium (Hannan, 1957; Moulin and Vial, 1978).[4] We formally contrast BCCEs with BCEs in Appendix C. If $\mathcal{G}(\lambda)$ is a game with complete information, BCCEs reduce to coarse correlated equilibria.

3.2 Asymptotic No Regret and Static Equilibria

We define the notion of an empirical distribution of actions, signals, and states from $\mathcal{G}^\infty$.

Definition 4 (Empirical Distribution)**.** *The empirical distribution of actions, signals, and states from $\mathcal{G}^\infty$ at the end of period N is denoted by $Z_N \in \Delta(A \times T \times \Theta)$, where*

$$Z_N(a, t, \theta) \coloneqq \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}_{\{a\}}(a_n)\mathbb{1}_{\{t\}}(t_n)\mathbb{1}_{\{\theta\}}(\theta_n).$$

$Z_N(a, t, \theta)$ is the empirical frequency of the action-signal-state profile (a, t, θ) in the first N periods. Note that q_N defined in (1) is the marginal on A of the empirical distribution Z_N for all positive integers N. That is, $q_N(a) = \sum_{t,\theta} Z_N(a, t, \theta)$ for all $a \in A$.

[4]In recent work, Brooks, Du, and Zhang (2024) and Zhang (2024) use the BCCE notion to construct robust bounds on economic outcomes.

The following theorem shows that a sequence of actions, signals, and states from $\mathcal{G}^\infty$ has ANR if and only if the sequence of empirical distributions of actions, signals, and states converges almost surely to the set of BCCEs of $\mathcal{G}(\lambda_0)$.

Theorem 1 (Convergence under ANR)**.** *The sequence $(a_n, t_n, \theta_n)_{n=1}^\infty$ from $\mathcal{G}^\infty$ has ANR if and only if the sequence of empirical distributions $(Z_N)_{N=1}^\infty$ converges almost surely to $E(\lambda_0)$.*

We next define restrictions on the set of actions implied by BCCE.

Definition 5 (BCCE Prediction)**.** *The probability distribution $q \in \Delta(A)$ is a BCCE prediction of $\mathcal{G}(\lambda)$ if there exists $\nu \in E(\lambda)$ such that*

$$q(a) = \sum_{t,\theta} \nu(a, t, \theta) \quad \textit{for all } a \in A.$$

We denote by $Q(\lambda)$ the set of BCCE predictions of $\mathcal{G}(\lambda)$.

The following result is an immediate corollary of Theorem 1.

Corollary 1 (Convergence of Empirical Distribution of Actions)**.** *If the sequence $(a_n, t_n, \theta_n)_{n=1}^\infty$ from $\mathcal{G}^\infty$ has ANR, then the sequence of empirical distributions of actions $(q_N)_{N=1}^\infty$ converges almost surely to $Q(\lambda_0)$.*

Convergence Notion. We clarify the convergence notion of $(q_N)_{N=1}^\infty$ to $Q(\lambda_0)$ in Corollary 1. Similar remarks apply to the convergence of $(Z_N)_{N=1}^\infty$ to $E(\lambda_0)$ in Theorem 1. For this purpose, we introduce the notion of an ε-BCCE prediction.

Definition 6 (ε-BCCE Prediction)**.** *Let $\varepsilon \geq 0$. The set of Bayes coarse correlated ε-equilibrium predictions of $\mathcal{G}(\lambda)$ is*

$$Q(\varepsilon; \lambda) := \{q \in \Delta(A) : d(q, Q(\lambda)) \leq \varepsilon\},$$

where, for the Euclidean norm $\|\cdot\|$, $d(q_N, Q(\lambda)) := \inf_{q \in Q(\lambda)} \|q_N - q\|$.

The almost sure convergence of $(q_N)_{N=1}^\infty$ to $Q(\lambda_0)$ in Corollary 1 means that

$$\mathbb{P}\Big(\lim_{N \to \infty} d(q_N, Q(\lambda_0)) = 0 \Big) = 1. \tag{2}$$

That is, the following statement holds almost surely: for all $\varepsilon > 0$, there exists $N_\varepsilon \in \mathbb{N}$ such that $q_N \in Q(\lambda_0; \varepsilon)$ for all $N \geq N_\varepsilon$. The a.s. convergence of $(q_N)_{N=1}^\infty$ is to $Q(\lambda_0)$, not necessarily to a point in that set. Moreover, whereas the empirical distribution of actions eventually becomes a BCCE prediction, there is no guarantee that play does so.

Discussion. Although only instrumental in this paper, Theorem 1 is of independent interest: it provides dynamic foundations for the static notions of BCE and BCCE (we discuss convergence to the set of BCEs in Appendix C). The theorem generalizes to games with incomplete information earlier work on the dynamic foundations for (coarse) correlated equilibrium in games with complete information (e.g., Foster and Vohra, 1997; Hart and Mas-Colell, 2000).

Existing work in computer science (Hartline, Syrgkanis, and Tardos, 2015; Caragiannis, Kaklamanis, Kanellopoulos, Kyropoulou, Lucier, Paes Leme, and Tardos, 2015) studies price-of-anarchy and efficiency properties of regret minimization in incomplete information games. Instead, our focus is on connecting regret minimization to an inference problem. Moreover, Hartline et al. (2015) study an independent private values setting in which private information is independent across agents and time, and Caragiannis et al. (2015) allow for correlated valuations in generalized second-price auctions. In contrast, we work with general games, allowing for private information correlated across agents and time and for common values.[5]

We highlight a central difference between our setup and that in Hartline et al. (2015), which gives rise to different convergence results. In Hartline et al. (2015), each agent's actions are conditionally independent of other agents' signals, given the agent's signal itself. In contrast, in our environment, agents may receive further signals informative about the payoff state; such signals may induce additional correlation across actions, which breaks conditional independence. Because of this contrast, our dynamics converge to the set of BCCEs—a superset of the limit set characterized by Hartline et al. (2015).

3.3 Illustration: A Two-Seller Pricing Game

Consider again our running illustration: the two-seller pricing game. According to Theorem 1, a sequence of prices and marginal costs $(p_n, t_n)_{n=1}^{\infty}$ from $\mathcal{G}^{\infty}$ has ANR if and only if, as $N \to \infty$, the sequence of empirical distributions of prices and marginal costs $(Z_N)_{N=1}^{\infty}$, where $Z_N(p,t) := \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}_{\{p\}}(p_n)\mathbb{1}_{\{t\}}(t_n)$, satisfies:

1. Consistency for $\mathcal{G}(\lambda_0)$: for all t, we have $\lim_{N\to\infty}\sum_p Z_N(p,t) = \psi(t;\lambda_0)$ a.s.
2. Coarse obedience for $\mathcal{G}(\lambda_0)$: for all i and t_i, we have

$$\limsup_{N\to\infty} \sum_{p,t_{-i}} \left[s(p_i', p_{-i};\lambda_0)(p_i' - t_i) - s(p;\lambda_0)(p_i - t_i)\right] Z_N(p,(t_i,t_{-i})) \leq 0 \quad \text{for all } p_i' \text{ a.s.}$$

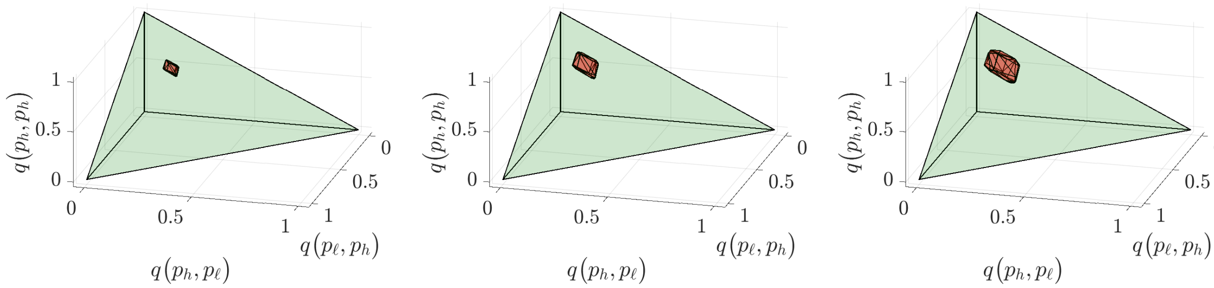
To illustrate the restrictions on actions implied by BCCE, we construct $Q(\lambda_0;\varepsilon)$ as an interpretable neighborhood of $Q(\lambda_0)$. In particular, we relax the coarse obedience constraints of each seller for each of his marginal costs by $\varepsilon = \tilde{\varepsilon}\kappa$, where $\tilde{\varepsilon}$ rescales the constant κ.[6] We

[5]For the analysis of settings with private information correlated over time, see Appendix B.

[6]This procedure is described in detail in Appendix E.1.

choose κ to capture the scale of payoffs (via the maximum payoff difference across actions over sellers and their marginal costs) and the cardinality of each seller's set of actions. We plot $Q(\lambda_0; \varepsilon)$ in Figure 2 for three values of ε; as expected, the set grows with ε.

Figure 2: Sets of ε-BCCE Predictions.

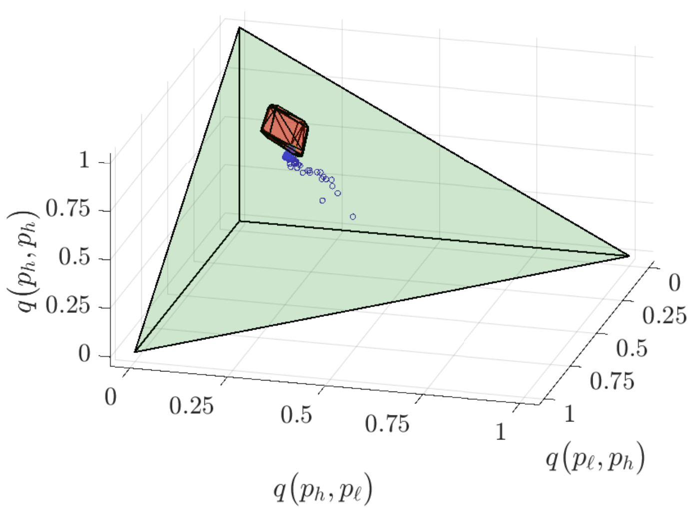


(a) $\varepsilon = 0.025\kappa$. (b) $\varepsilon = 0.05\kappa$. (c) $\varepsilon = 0.1\kappa$.

The light-red convex sets correspond to $Q(\lambda_0; \varepsilon)$ for $\varepsilon = 0.025\kappa$, $\varepsilon = 0.05\kappa$, and $\varepsilon = 0.1\kappa$.

Figure 3 represents the empirical distribution of prices corresponding to $N = 10,000$ iterations of regret matching. Each blue dot is a snapshot of such empirical distribution at a point along the path. The light-red convex set represents $Q(\lambda_0; \varepsilon)$ for $\varepsilon = 0.05\kappa$. The convergence of the empirical distribution of prices to $Q(\lambda_0; \varepsilon)$ illustrates Corollary 1.

Figure 3: Convergence under Asymptotic No Regret.



We represent a path of the empirical distribution of prices $(q_K)_{K=1}^N$ generated by $N = 10,000$ iterations of regret matching. Blue dots correspond to the empirical distribution of prices at different points, each 100 periods apart, along the path. The light-red convex set corresponds to $Q(\lambda_0; \varepsilon)$ for $\varepsilon = 0.05\kappa$.

4 Econometrics

In this section, we build on the theoretical results we established in Section 3 to develop an empirical strategy for (partially) identifying and estimating λ_0 under Assumption 1.

4.1 Identification

Corollary 1 only ensures that $(q_N)_{N=1}^{\infty}$, the sequence of empirical distributions of actions observed by the researcher, converges almost surely to $Q(\lambda_0)$, the set of BCCE predictions of the stage game $\mathcal{G}(\lambda_0)$, not necessarily to a specific element of that set. Hence, the convergence of $(q_N)_{N=1}^{\infty}$ to a single limiting "population" distribution of the observable actions, $q_0 \in \Delta(A)$—which is straightforward when the repetition of independent and identical experiments generates the data—is not guaranteed under ANR.

This feature of the data-generating process has implications for how we define the identified set—condition (2) may also hold for some other $\lambda \in \Lambda$, potentially leading to a loss of point identification for λ_0 and the need to characterize an identified set. Following Manski (2003), an identified set is usually defined as the set of parameters compatible with the single probability measure asymptotically revealed by the sampling process. Our empirical environment, however, departs from standard identification analysis: the researcher does not (asymptotically) observe a single distribution of the observable $q_0 \in \Delta(A)$ upon which to characterize the identified set of parameters.

We address this issue by formulating a definition of the identified set in the space of action sequences. By Theorem 1 and Corollary 1, although agents' interaction is dynamic and we do not maintain that data are generated by equilibrium play, the static model of BCCE provides sharp and tractable restrictions to recover valid bounds on the structural parameters in our setting. Formally, we have the following definition of the identified set.

Definition 7 (Identified Set)**.** *Under Assumption 1, the identified set of parameters is*

$$\Lambda^* := \Big\{\lambda \in \Lambda : \mathbb{P}\Big(\lim_{N\to\infty} d(q_N, Q(\lambda)) = 0\Big) = 1\Big\}.$$

The identified set Λ^* consists of all parameters $\lambda \in \Lambda$ that are compatible with the observed empirical distribution of actions as a BCCE prediction of the stage game $\mathcal{G}(\lambda)$ as the sample size grows large. Clearly, $\lambda_0 \in \Lambda^*$.

4.2 Estimation

We now introduce the notion of an ε-BCCE estimator of the identified set Λ^*. The estimator uses the observed empirical distribution of actions q_N to recover the structural parameters. In particular, for all $\varepsilon > 0$, the ε-BCCE estimator, denoted by $\widehat{\Lambda}_N(\varepsilon)$, consists of all parameters $\lambda \in \Lambda$ that are compatible with q_N as an ε-BCCE prediction of the stage game.

Definition 8 (ε-BCCE Estimator)**.** *Let $\varepsilon > 0$ and q_N be the observed empirical distribution of N action profiles. The ε-BCCE estimator of the identified set Λ^* is*

$$\widehat{\Lambda}_N(\varepsilon) := \{\lambda \in \Lambda : q_N \in Q(\varepsilon; \lambda)\}.$$

Since we have an ε-BCCE estimator for all $\varepsilon > 0$, Definition 8 practically defines a class of estimators, one for each $\varepsilon > 0$. The following theorem shows that, for all $\varepsilon > 0$ and $\lambda \in \Lambda^*$, the parameters λ are contained in the ε-BCCE estimator $\widehat{\Lambda}_N(\varepsilon)$ for all sufficiently large N.

Theorem 2 (Almost Sure Coverage of ε-BCCE Estimators)**.** *Under Assumption 1, for all $\varepsilon > 0$ and $\lambda \in \Lambda^*$,*

$$\mathbb{P}\Big(\lambda \in \liminf_{N\to\infty} \widehat{\Lambda}_N(\varepsilon)\Big) = 1.$$

Theorem 2 establishes that, for all $\varepsilon > 0$, the static equilibrium notion of ε-BCCE provides valid restrictions for the estimation of dynamic interactions that satisfy ANR. In particular, the restrictions implied by the ε-BCCE notion lead to estimating a set that contains almost surely the true parameters λ_0 as the sample size grows large.

The strong (a.s.) nature of the coverage result in Theorem 2 derives from the theoretical convergence property of the data under ANR. Despite data not being generated by repeated identical experiments, we bound parameters without statistical assumptions on the sampling process on top of the ANR assumption on behavior.

As reflected in the definitions of the identified set and estimator, we focus on recovering structural parameters from data, as opposed to Nekipelov et al. (2015), who also recover the level of ε consistent with the data. The reason is that jointly identifying (λ, ε) can be computationally demanding outside their second-price auction environment and, more generally, may result in uninformative sets. Instead, as we formalize in the next section, we leverage discrepancies from perfect regret minimization and the convergence rates of regret-minimizing AIs to conduct inference in finite samples.

4.3 Confidence Regions and Coverage

Our definitions of identified set Λ^* and ε-BCCE estimator $\widehat{\Lambda}_N(\varepsilon)$ rely only on the ANR assumption, and so does the almost sure coverage property of ε-BCCE estimators in Theorem 2. A natural next question is how the researcher should choose ε to ensure that desired coverage properties are satisfied in finite samples. The lack of convergence of $(q_N)_{N=1}^{\infty}$ to a single limiting distribution of actions highlighted for identification analysis in Section 4.1 also implies that our empirical environment departs from standard estimation and inferential analysis.[7]

We address these issues by constructing the set of ε-BCCE predictions $Q(\cdot;\varepsilon)$ so that the corresponding ε-BCCE estimator $\widehat{\Lambda}_N(\varepsilon)$ becomes a confidence region for all $\lambda \in \Lambda^*$, ensuring uniform validity. We do so by further exploiting the theoretical convergence properties of ANR algorithms.

To this purpose, we need to take a stance about the three features we discuss in Section 2.2.1: the environment in which agents use ANR algorithms, the informational feedback to the algorithms, and the algorithms' objective (beyond satisfying ANR). Which assumptions are more appropriate depends on the specific application. The most robust approach is maintaining the adversarial bandit environment with partial feedback for all agents. Alternatively, the researcher can strengthen these assumptions, e.g., by maintaining that each agent adopts a specific ANR algorithm (possibly different across agents) or that agents coordinate on a particular algorithm or class. Under these assumptions, the researcher can use theoretical results on the convergence of ANR algorithms to construct a confidence region for the identified set.[8]

To construct uniformly valid confidence regions, we first introduce some auxiliary concepts. Let $\underline{\varepsilon} := (\varepsilon(i,t_i))_{i\in\mathcal{I},t_i\in T_i}$, where $\varepsilon(i,t_i) \geq 0$ for all $i \in \mathcal{I}$ and $t_i \in T_i$. That is, $\underline{\varepsilon}$ specifies a non-negative real number for all i and t_i. We next define the notion of $\underline{\varepsilon}$-BCCE prediction.

Definition 9 ($\underline{\varepsilon}$-BCCE and $\underline{\varepsilon}$-BCCE Prediction)**.** *Let $\underline{\varepsilon} := (\varepsilon(i,t_i))_{i\in\mathcal{I},t_i\in T_i}$, where $\varepsilon(i,t_i) \geq 0$ for all $i \in \mathcal{I}$ and $t_i \in T_i$. The probability distribution $\nu \in \Delta(A \times T \times \Theta)$ is a Bayes coarse correlated $\underline{\varepsilon}$-equilibrium ($\underline{\varepsilon}$-BCCE) of $\mathcal{G}(\lambda)$ if:*

1. *ν is consistent for $\mathcal{G}(\lambda)$ (see Definition 3).*
2. *ν is coarsely $\underline{\varepsilon}$-obedient for $\mathcal{G}(\lambda)$; that is, for all $i \in \mathcal{I}$ and $t_i \in T_i$, we have*

$$\sum_{a,t_{-i},\theta} \left[u_i\big((a_i', a_{-i}), \theta; \lambda\big) - u_i(a,\theta;\lambda)\right]\nu(a,(t_i,t_{-i}),\theta) \leq \varepsilon(i,t_i) \quad \textit{for all } a_i' \in A_i.$$

We denote by $Q(\underline{\varepsilon};\lambda) \subseteq \Delta(A)$ the set of $\underline{\varepsilon}$-BCCE predictions of $\mathcal{G}(\lambda)$.

An $\underline{\varepsilon}$-BCCE of $\mathcal{G}(\lambda)$ is obtained by relaxing each coarse obedience constraint in the definition of a BCCE (see Definition 3) by $\varepsilon(i,t_i)$. The set of $\underline{\varepsilon}$-BCCE predictions of $\mathcal{G}(\lambda)$ is the set of marginals on the set of action profiles A of $\underline{\varepsilon}$-BCCE distributions.

Definition 10 ($\underline{\varepsilon}$-BCCE Confidence Region)**.** *Let $\underline{\varepsilon} := (\varepsilon(i,t_i))_{i\in\mathcal{I},t_i\in T_i}$, where $\varepsilon(i,t_i) \geq 0$ for all $i \in \mathcal{I}$ and $t_i \in T_i$, and q_N be the observed empirical distribution of N action profiles.*

[7] These features are reminiscent of Epstein, Kaido, and Seo (2016), who study inference in complete information static games without observing i.i.d. draws from a single $q_0 \in \Delta(A)$.

[8] The structure of the empirical model should be invariant to any assumption beyond ANR the researcher may want to impose. For an example of how to tailor such additional assumptions to an empirical model without affecting the identified set, we refer to our application in Section 5.

The $\underline{\boldsymbol{\varepsilon}}$-BCCE confidence region is

$$\widehat{\Lambda}_N(\underline{\boldsymbol{\varepsilon}}) := \{\lambda \in \Lambda : q_N \in Q(\underline{\boldsymbol{\varepsilon}}; \lambda)\}.$$

The following theorem shows that, by postulating assumptions on 1–3 in Section 2.2.1 and appropriately specifying $\varepsilon(i, t_i)$ for all $i \in \mathcal{I}$ and $t_i \in T_i$, the set $\widehat{\Lambda}_N(\underline{\boldsymbol{\varepsilon}})$ becomes a uniformly valid conservative confidence region for all $\lambda \in \Lambda^*$ at any desired confidence level. To state the theorem, we need one last bit of notation: for all $i \in \mathcal{I}$, $t_i \in T_i$, and $\lambda \in \Lambda$, let

$$K(i, t_i; \lambda) := \max_{a,a'} |u_i(a, t_i; \lambda) - u_i(a', t_i; \lambda)|,$$

be the maximum payoff difference across action profiles for agent i with signal t_i, and let

$$\phi(t_i; \lambda) := \mathbb{P}(t_{i,n} = t_i) = \sum_{\theta} \sum_{t_{-i}} \pi(t_i, t_{-i} \,|\, \theta; \lambda)\psi(\theta; \lambda).$$

be the probability of signal t_i.

Theorem 3 (Uniformly Valid $\underline{\boldsymbol{\varepsilon}}$-BCCE Confidence Region)**.** *Suppose Assumption 1 holds. For all agents $i \in \mathcal{I}$, assume: 1. adversarial bandit environment; 2. bandit feedback; 3. the algorithm optimizes the convergence rate of the expected regret to 0 given 1. and 2. Fix $\alpha \in (0, 1)$ and $N \in \mathbb{N}$. For all $i \in \mathcal{I}$, $t_i \in T_i$, and $\lambda \in \Lambda$, let*

$$\varepsilon_{\alpha,N}(i, t_i; \lambda) := \frac{\left[K(i, t_i; \lambda)\sqrt{\ln |A_i|}\sqrt{\phi(t_i; \lambda)}\right]\left[\sum_{i\in\mathcal{I}} |T_i|\right]}{\alpha\sqrt{N}}.$$

Then, $\widehat{\Lambda}_N(\underline{\boldsymbol{\varepsilon}}_{\alpha,N}) := \left\{\lambda \in \Lambda : q_N \in Q(\underline{\boldsymbol{\varepsilon}}_{\alpha,N}(\lambda); \lambda)\right\}$ is a uniformly valid conservative confidence region for all $\lambda \in \Lambda^*$; that is,

$$\inf_{\lambda\in\Lambda^*} \mathbb{P}\Big(\lambda \in \widehat{\Lambda}_N(\underline{\boldsymbol{\varepsilon}}_{\alpha,N})\Big) \geq 1 - \alpha.$$

By postulating different assumptions on 1–3 above than those in Theorem 3 and specifying $\underline{\boldsymbol{\varepsilon}}_{\alpha,N}$ accordingly, one can construct a uniformly valid conservative confidence region for the identified set analogously, under these alternative assumptions.

The bounds in Theorem 3 (or those obtained under different assumptions on 1–3 above) are conservative for several reasons. First, the confidence region is obtained under worst-case bounds on the convergence properties of ANR algorithms. Worst-case bounds are mathematical guarantees that must hold under the most unfavorable sequence of events, thus leading to the highest possible regret. Second, to obtain the confidence region in Theorem 3, we use results on the convergence rate of the expected regret to 0 together with Markov's inequality. Since Markov's inequality uses no information about the distribution and also provides worst-case bounds, this operation adds another layer of conservativeness.

Many simulation studies document that the practical performance of ANR algorithms is significantly better than worst-case theoretical bounds and that the empirical distribution of realized regrets often concentrates around its expectation. We further explore this observation in our simulations.

4.4 Computation

To compute the ε-BCCE estimator $\widehat{\Lambda}_N(\varepsilon)$ and the $\underline{\boldsymbol{\varepsilon}}$-BCCE confidence region $\widehat{\Lambda}_N(\underline{\boldsymbol{\varepsilon}})$, we follow Magnolfi and Roncoroni (2023) and first transform these sets as zero-level sets of a function of parameters. Denote by $b^\intercal$ the transpose of $b \in \mathbb{R}^{|A|}$. The following theorem holds.

Theorem 4 (Computation of Estimator and Confidence Region)**.** *For all $\varepsilon > 0$, consider the function $g(\cdot; q_N, \varepsilon)\colon \Lambda \to \mathbb{R}$ defined as*

$$g(\lambda; q_N, \varepsilon) := \max_{b \in \mathbb{R}^{|A|}} \min_{\substack{q \in \mathbb{R}_+^{|A|} \\ \nu \in \mathbb{R}_+^{|A||T||\Theta|}}} b^\intercal (q_N - q) \tag{P1}$$

subject to

$$\begin{aligned} b^\intercal b - 1 &\leq 0, \\ q(a) - \sum_{t,\theta} \nu(a,t,\theta) &= 0 \quad \textit{for all } a \in A, \\ \sum_a \nu(a,t,\theta) - \pi(t \mid \theta; \lambda)\psi(\theta; \lambda) &= 0 \quad \textit{for all } t \in T, \\ \sum_{a, t_{-i}, \theta} \big[u_i\big((a_i', a_{-i}), \theta; \lambda\big) - u_i(a, \theta; \lambda)\big]\nu(a, (t_i, t_{-i}), \theta) - \varepsilon &\leq 0 \quad \textit{for all } i \in \mathcal{I}, a_i' \in A_i, t_i \in T_i. \end{aligned}$$

Then, the ε-BCCE estimator $\widehat{\Lambda}_N(\varepsilon)$ has the following characterization:

$$\widehat{\Lambda}_N(\varepsilon) = \{\lambda \in \Lambda : g(\lambda; q_N, \varepsilon) = 0\}.$$

An analogous characterization holds for the $\underline{\boldsymbol{\varepsilon}}$-BCCE confidence region $\widehat{\Lambda}_N(\underline{\boldsymbol{\varepsilon}})$.

According to Theorem 4, computing the set estimator $\widehat{\Lambda}_N(\varepsilon)$ amounts to evaluating the function $g(\cdot; q_N, \varepsilon)$ on a finite grid over Λ. In turn, finding $g(\cdot; q_N, \varepsilon)$ entails solving the program (P1), a max-min program with one convex and multiple linear constraints. Together, the second, third, and fourth constraints are equivalent to $q \in Q(\lambda; \varepsilon)$. In particular, the third and fourth constraints correspond to the ε-BCCE restrictions on ν (consistency and coarse ε-obedience), and the second constraint requires q to be the ε-BCCE prediction corresponding to ν. The first constraint is equivalent to $b \in B^{|A|}$, where $B^{|A|} := \{b \in \mathbb{R}^{|A|} : b^\intercal b \leq 1\}$ is the closed unit ball centered at $0_{|A|} \in \mathbb{R}^{|A|}$. This latter constraint corresponds to the support-function characterization of the non-empty, closed, and convex set $Q(\varepsilon; \lambda)$ (for further details, see Appendix A.4).

Our definition of g is general: different formulations of the basic game will only affect the dimensions of the problem and how λ enters $u_i(\cdot;\lambda)$, $\pi(\cdot\,|\,\cdot;\lambda)$, and $\psi(\cdot;\lambda)$. The computation of $g(\cdot;q_N,\varepsilon)$ can be further simplified by replacing the inner constrained minimization problem in program (P1) by its dual, which consists of a linear constrained maximization problem. Therefore, we can check whether a given value of λ belongs to the estimator $\widehat{\Lambda}_N(\varepsilon)$ by solving a single linear constrained maximization problem. An analogous characterization, with the obvious changes, holds for the confidence region $\widehat{\Lambda}_N(\underline{\boldsymbol{\varepsilon}}_{\alpha,N})$. Appendix E includes details on the application of duality and the full formulation of the program.

Computing our estimator or confidence region involves a grid search over Λ. If parameters are multidimensional, this search is exponential in discretization width. This observation applies to many models in this class (see, e.g., Molinari, 2020, for a discussion). In our simulations and empirical application, such a grid search is computationally manageable. If the researcher adopts a non-parametric specification of the unobserved payoff heterogeneity, i.e., $\lambda=\psi$, she can avoid the grid search by adopting the method in Syrgkanis et al. (2021) to compute counterfactual bounds in polynomial time.[9]

4.5 Illustration: A Two-Seller Pricing Game

We illustrate the estimation strategy in the context of our running illustration: the two-seller pricing game. We generate data under a parametric and numerical specification similar to that in Section 3.3 but with a richer set of actions. Sellers set prices by choosing among five equally-spaced values in $[4,8]$, and the demand parameter (known to the researcher) is $\eta=-1/3$. Marginal costs are i.i.d. across sellers and over time according to a discretized truncated Normal distribution with $\mu_0=3$, $\sigma_0=1$, taking values over five equally spaced values between $\underline{t}=0$ and $\bar{t}=6$. Sellers set prices using regret matching, which we iterate for N periods to obtain the empirical distributions of prices q_N observed by the researcher. We use these data to estimate the structural parameters $\lambda_0=(\mu_0,\sigma_0)$.

To construct a confidence region for the identified set, we choose $\underline{\boldsymbol{\varepsilon}}_{\alpha,N}(\lambda)$ to guarantee a regret lower than $\varepsilon_{\alpha,N}(i,t_i;\lambda)$ with probability 0.95 for all $i\in\mathcal{I}$, $t_i\in T_i$, and $\lambda\in\Lambda^*$ under the assumptions in Theorem 3. This approach is robust because it relies on worst-case bounds in an adversarial environment with bandit feedback. On the flip side, these conservative assumptions require large sample sizes for informative inference. Large samples, however, are routine in most algorithmic environments where, e.g., pricing or bidding AIs make high-frequency decisions at scale.

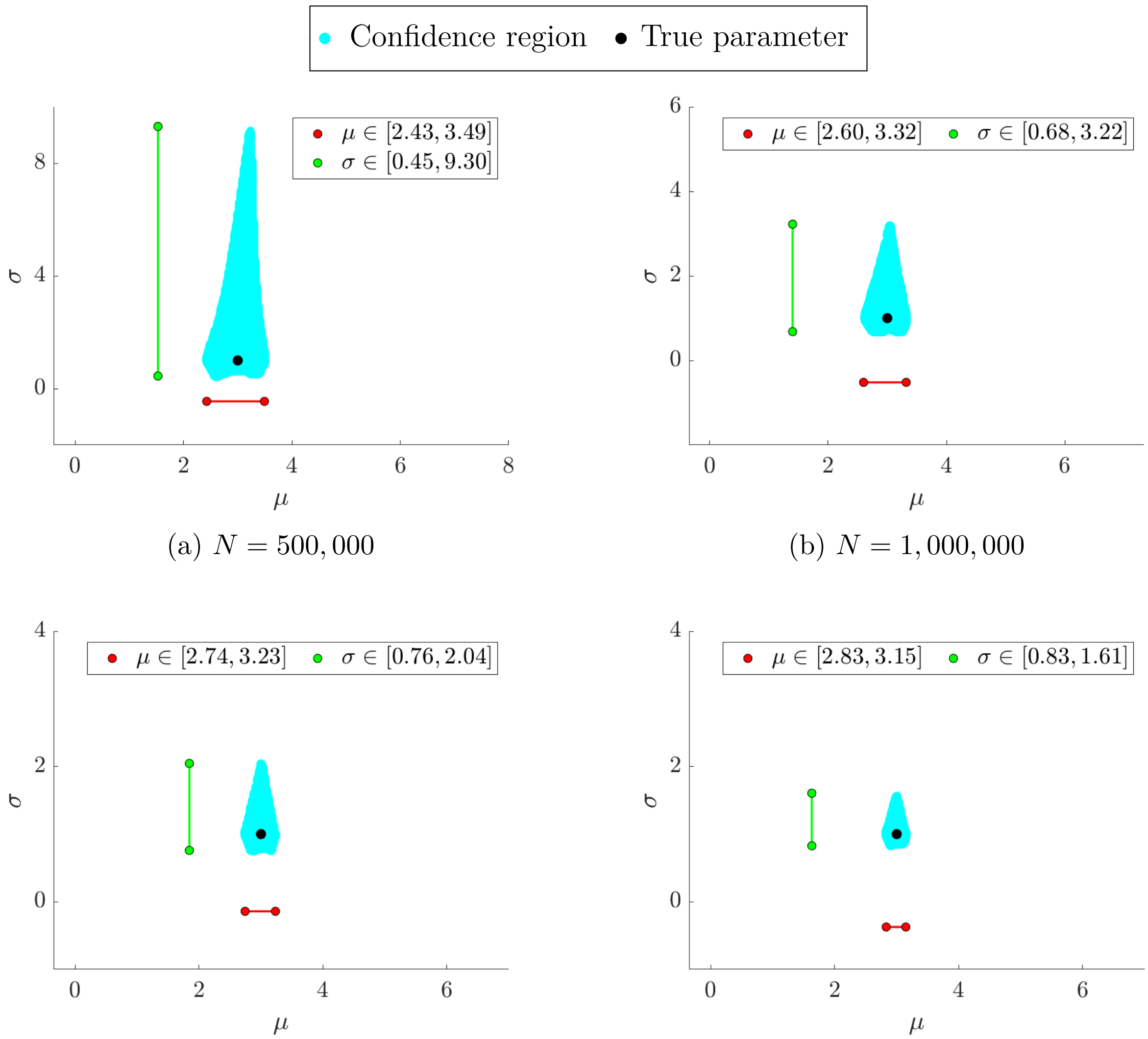
We illustrate our simulation results in Figure 4, which shows confidence regions for four

[9] Following Theorem 5 in Syrgkanis et al. (2021), for a counterfactual $F(a,t,\theta)$, the researcher formulates the program as $\min/\max_{\nu\in\mathbb{R}_+^{|A||T||\Theta|}}\sum_{a,t}F(a,t,\theta)\nu(a,t,\theta)$ subject to the linear constraints in program (P1).

different sample sizes N. The parameter μ is estimated with reasonable precision even with $500,000$ observations (panel (a)). As the number of observations increases, the confidence region projection for μ shrinks, reaching $[2.83, 3.15]$ with 4 million observations (panel (d)).

Estimating σ is more challenging. The confidence region with $500,000$ observations is not very informative, with a large upper bound of 9.30 (recall that $\sigma_0 = 1$). Larger sample sizes, however, yield more informative confidence regions even under our conservative assumptions. With 4 million observations, the confidence region projection for σ narrows to $[0.83, 1.61]$.

Figure 4: Confidence Regions for Different Sample Sizes.



(a) $N = 500,000$

(b) $N = 1,000,000$

(c) $N = 2,000,000$

(d) $N = 4,000,000$

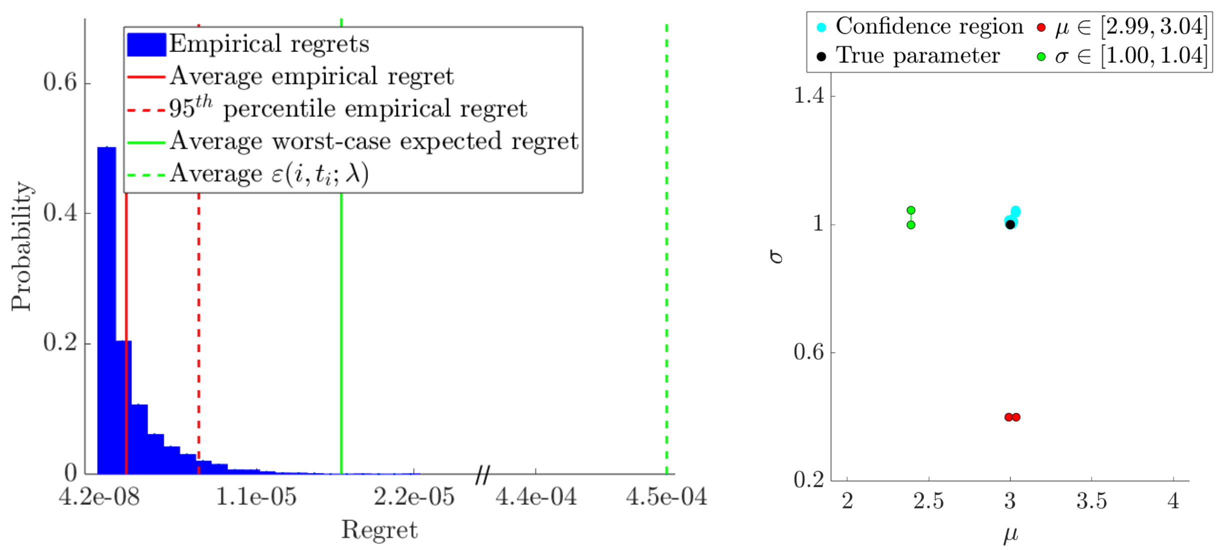
We represent in blue confidence regions in the (μ, σ)-parameter space for different sample sizes N. Projections for μ and σ are reported in red and green. We compute each confidence region by specifying $\underline{\varepsilon}_{\alpha,N}(\lambda)$ to guarantee for each seller i and marginal cost t_i a regret lower than $\varepsilon_{\alpha,N}(i, t_i; \lambda)$ with probability 0.95 for each λ in the identified set. Computational details are in Appendix E.2.

Figure 4 suggests that inference under ANR with worst-case assumptions requires large (though common in many algorithmic environments) sample sizes. Following our discussion in Section 4.3, however, two additional remarks are in order. First, the empirical environment may suggest assumptions that enable the researcher to conduct inference with smaller samples. This is the case for our empirical application in Section 5.

Second, the environment is often less unfavorable or adversarial than under our assumptions, making our inference very conservative. This is evident in our simulations, where both agents use a regret-matching algorithm and observe each other's actions, leading to a much faster convergence of expected regrets to 0 than prescribed by worst-case bounds.

To illustrate the latter point, we run 500 independent instances of the pricing game described above and construct the distribution of empirical regrets in period $N = 100,000$ across sellers and marginal costs. We plot this distribution in Figure 5, panel (a), comparing it with the average worst-case expected regret and the corresponding $\varepsilon(i, t_i; \lambda_0)$. Examining the distribution of empirical regrets, we observe that these are, on average, about an order of magnitude smaller than the worst-case expected regrets and two orders of magnitude smaller than $\varepsilon(i, t_i; \lambda_0)$. This result is consistent with well-documented evidence in computer science (see, e.g., Farina, Grand-Clément, Kroer, Lee, and Luo, 2024).

Figure 5: Empirical Regrets and Corresponding Confidence Region for $N = 100,000$.



(a) Distribution of Empirical Regrets

(b) Confidence Region

Panel (a) represents the distribution of empirical regrets for a simulation exercise where data were simulated 500 times with regret matching for $N = 100,000$ periods. The solid lines represent average (across sellers and types) empirical and worst-case regrets (respectively in red and green). The dashed lines represent the 95^{th} percentile of empirical regrets and the average $\varepsilon(i, t_i; \lambda)$ corresponding to worst-case bounds (respectively in red and green). Panel (b) represents the confidence region for (μ, σ) obtained using average empirical regrets to construct bounds. Computational details are in Appendix E.2.

To assess the impact on inference, we construct the confidence region containing the true parameter with a probability of at least 0.95 according to the empirical distribution of regrets in our simulations. The resulting confidence region is plotted in Figure 5, panel (b). Despite using a much smaller sample size ($N = 100,000$) than in the previous simulation exercise, the confidence region constructed with empirical regrets is significantly smaller, pinpointing the true parameters with high precision. This exercise suggests that adopting bounds on regret that are significantly smaller than those implied by worst-case bounds may be justified in applications, especially if motivated by simulation studies and knowledge of the environment. We further exploit these ideas in the context of our empirical application.

5 Application: Pricing in an Online Platform

We apply our method to analyze sellers' pricing behavior on Swappa, a decentralized marketplace for used smartphones with around \$100 million in annual transactions.[10] Swappa provides a good empirical setting for our study due to its complex pricing environment: sellers navigate a rapidly changing competitive landscape, adjusting their pricing strategies for a diverse set of devices. This scenario mirrors the challenges faced by individual sellers on other decentralized platforms and marketplaces. The nature of pricing decisions on the platform makes Swappa a prime candidate for implementing AI or AI-assisted pricing decisions.

In this context, we employ our method to estimate the distribution of sellers' marginal costs, a fundamental parameter for addressing counterfactual questions, e.g., on platform design. We use the primitive to quantify markups for sellers on Swappa and find that these are lower than for competing centralized platforms that resell used electronics.

The following subsections lay out our approach: first, we describe the empirical setting and data; next, we present our empirical model; finally, we analyze the estimation results.

5.1 Empirical Setting and Data

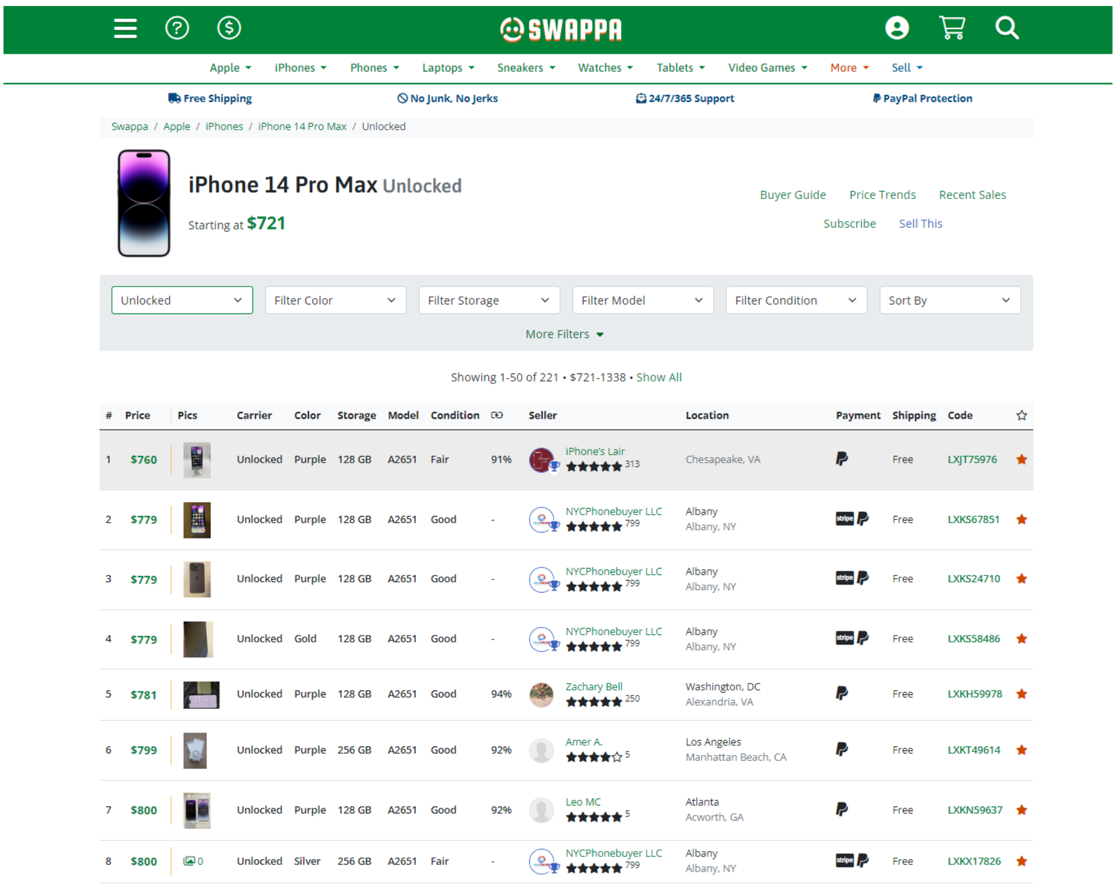
Swappa is a decentralized digital marketplace for used smartphones. The platform has a thorough approval process to ensure all listed devices are activation-ready. Sellers incur a nominal listing fee. Buyers pay a fee included in the purchase price upon transaction completion. While some sellers allow for international shipping, the platform and most users are US-based. The shopping experience on Swappa is straightforward: buyers first select the desired device generation (e.g., iPhone 14 series) and then browse listings within that category. For visual reference, Figure 6 provides a website screenshot that shows the user interface.

We collected data from Swappa by scraping its website for six consecutive months, from

[10] See `https://swappa.com/about`.

July to December 2023. While Swappa offers a variety of electronic devices, we focused our data collection on the most popular iPhone models, as these represent the most active markets on the platform. For each listing, each corresponding to a device on sale, we gathered daily information, including price, seller characteristics, product specifications, and a unique identifier. We use identifiers to track the final sale prices and dates of completed transactions. Over the sample period, we collected 259,102 device-day observations from 4,980 sellers.

Figure 6: A Screenshot from the Swappa Website.



Swappa webpage that users see when selecting a device model (in this case, iPhone 14 Pro Max).

On Swappa, sellers range from one-time users listing a single device to professional merchants who regularly post multiple devices. While we collect data for all sellers in the marketplace and use the entire dataset to learn about demand (i.e., the probability of selling a device given the price), our analysis focuses on experienced sellers who make multiple decisions over time. Therefore, we narrowed our focus to the most prolific sellers, applying several criteria. We considered high listing volume, with the top sellers having more than 1,500 listings in our sample period. We also looked at consistent activity, considering sellers with active listings throughout the sample period. This filtering process yields two sets of sellers: a core group of 15 sellers, which sell the majority of devices on the platform, and a group of two top sellers, which are significantly larger than the other sellers on the platform and make many pricing decisions. The former core group has the largest impact on competitive conditions on the platform. The latter group includes the most experienced sellers, with more than 10,000 device-day observations: these sellers are the most likely to use AI tools. All the top-15 sellers are firms specialized in acquiring, refurbishing, and reselling used smartphones. For these firms, a device's marginal cost primarily comprises acquiring used devices and refurbishing expenses.

Swappa represents a small fraction of the broader market for used electronics. To provide context for pricing decisions on Swappa within its broader market, we collected data on "reference prices" from Decluttr, a US-based centralized platform for used smartphones.[11] Decluttr is representative of similar platforms that directly purchase devices from consumers, refurbish them, and resell them. Comparing Decluttr's prices with Swappa's transaction data provides insight into Swappa sellers' decision-making within a competitive landscape. Unlike Swappa's decentralized model, Decluttr operates as a centralized platform. Buyers on Decluttr select their desired device model (e.g., iPhone 12 Pro, 128GB, in mint state) and then choose from available listings. This structure offers a standardized comparison point for our analysis. Our data collection from Decluttr covers the same sample period of the Swappa dataset. We gathered information about price, date, and device model (including storage capacity, color, and condition). For each device and day pair in our Swappa dataset, we match the corresponding reference price for the device model from Decluttr data. Descriptive statistics for Swappa and Decluttr data are in Appendix F.1, and more details on Swappa sellers are in Appendix F.2.

5.2 Empirical Model

Actions, Periods, and Payoffs. For each device j each seller i has on sale on Swappa, we assume the seller makes a pricing decision every day d that the listing is active.[12] Hereafter, we refer to each device-day pair (j, d) as a different period n; that is, $n = (j, d)$.

We define the set of competitors $-i$ of each device-day (j, d) of seller i as all devices of the same generation as device j (e.g., iPhone 12) available on Swappa on day d. Seller i's payoff in period n is $\tilde{u}_i(p_{i,n}, p_{-i,n}, t_i) := \tilde{s}_i(p_{i,n}, p_{-i,n})(p_{i,n} - t_{i,n})$, where: $p_{i,n}$ and $p_{-i,n}$ are the prices of seller i and its competitors in period n; the function $\tilde{s}_i(p_{i,n}, p_{-i,n})$ maps price profiles into the probability of seller i making a sale in period n at those prices; $t_{i,n}$ is seller i's marginal cost in period n. This formulation creates distinct markets for each iPhone generation, thus

[11] See https://www.decluttr.com/.

[12] In Appendix F.3, we present evidence that the daily pricing assumption fits our environment well.

restricting the pool of competing devices on the platform.

Reference Pricing. To aggregate information across different device models, we address observed heterogeneity by modeling costs and pricing decisions as deviations from a model-specific reference price.[13] By doing so, we "homogenize" devices with varying characteristics such as storage capacity and condition. For each device-day pair $n = (j, d)$, we denote by χ_n the reference price for device j of model $m(j)$ on day d. We derive such reference prices from Decluttr data and make two assumptions.

First, we decompose marginal costs as $t_{i,n} = \chi_n + \zeta_{i,n}$, where $\zeta_{i,n} \in T_i$ is a seller-device-specific cost shock in period n. Hence, T_i is the finite set of seller i's marginal cost residuals. For each seller i, we assume marginal cost residuals $\zeta_{i,n}$ are drawn i.i.d. from the distribution $\psi_i(\cdot; \lambda_i) \in \Delta_{++}(T_i)$, parametrized by λ_i. Let $\lambda_{i,0}$ denote the true structural parameters in the data-generating process. The idiosyncratic cost component $\zeta_{i,n}$ captures acquisition cost or valuation specific to seller i. As sellers typically acquire devices below the retail price on other platforms, we expect $\zeta_{i,n}$ to be negative.

Second, we decompose prices as $p_{i,n} = \chi_n + \rho_{i,n}$, where $\rho_{i,n} \in A_i$ is the seller-chosen deviation from the reference price, or *pricing residual*. Hence, A_i is the finite set of seller i's actions. We rewrite seller i's payoff in period n as $u_i(\rho_{i,n}, \rho_{-i,n}, \zeta_{i,n}) := s_i(\rho_{i,n}, \rho_{-i,n})(\rho_{i,n} - \zeta_{i,n})$, where $s_i(\cdot)$ models sale probabilities for seller i. If $(\rho_{i,n}, \rho_{-i,n}) = (\rho_{i',n}, \rho_{-i',n})$, we assume that $s_i(\rho_{i,n}, \rho_{-i,n}) = s_{i'}(\rho_{i',n}, \rho_{-i',n})$, so that sale probabilities depend only on pricing residuals. Sale probabilities vary according to the discount or premium of a device on Swappa versus its reference price, as consumers may buy on other platforms if prices are too high. Appendix F.4 shows the variation in pricing residuals.

Our assumptions on reference pricing simplify the pricing problem by eliminating dependence on device characteristics and allow us to capture the fundamental incentive structure across different pricing problems.[14]

Aggregation of Competitors. The pricing game on Swappa is potentially large, with a fast-evolving set of many competing devices and sellers. To reduce the dimensionality of the game, we assume that seller i's payoff, through the function $s_i(\cdot)$, only depends on the aggregate state $\overline{\rho}_{-i,n} \in \Theta_i$ in period n, where Θ_i is some finite set. State $\overline{\rho}_{-i,n}$ captures prices of sellers other than i in period n for devices of the same generation. This modeling choice is in the spirit of the oblivious equilibrium approach (Weintraub, Benkard, and Van Roy, 2008) and of aggregative- and large-games approaches (e.g., Jensen, 2018; Gradwohl and Kalai, 2021).

[13] This choice reflects the economics of our application and allows us to learn from pricing decisions across different devices. Our general method, however, can account for observable heterogeneity (see Appendix D.1).

[14] Considering pricing residuals as multiplicative (instead of additive) deviations from reference prices would not achieve the same goal, as payoffs would still depend on device characteristics through the reference price.

Parametrization. We discretize pricing residuals by assigning the observations of $\rho_{i,n}$ to five bins, corresponding to equally-distant quantiles of the empirical distribution of $\rho_{i,n}$ across sellers and periods.[15] Hence, $A_i = A$ for all sellers i, with $|A| = 5$. We discretize $\overline{\rho}_{-i,n}$ similarly; hence, $\Theta_i = \Theta = A$ for all sellers i. We also assume that $\overline{\rho}_{-i,n}$ are i.i.d. with distribution $\psi \in \Delta_{++}(\Theta)$ that we estimate from the data, and that ζ_i and $\overline{\rho}_{-i}$ are independent.

With this discretization, we estimate the 25 possible values that function $s_i(\cdot)$ can take from the data. For all pairs $(\rho_i, \overline{\rho}_{-i})$, we compute $s_i(\rho_i, \overline{\rho}_{-i})$ as the average probability of selling a device across all observations with pricing residual $\rho_i = \rho$ and state $\overline{\rho}_{-i}$. Appendix F.4 shows our estimates of $s_i(\cdot)$, which follow an intuitive pattern.

We assume that each seller i's marginal cost shocks $\zeta_{i,n}$ are i.i.d. according to a truncated, discretized Normal distribution with seller-specific parameters μ_i and σ_i. The support of such distribution is common across sellers: $T_i = T$ for all sellers i, with $|T| = 5$. Let $\lambda_i \coloneqq (\mu_i, \sigma_i) \in \Lambda \coloneqq \mathbb{R} \times \mathbb{R}_+$, where μ_i determines the average level of cost residuals and σ_i captures the variability of seller i's marginal costs across periods. This specification allows for heterogeneous variability in marginal costs across periods (or device-day pairs) for each seller.

Empirical Game. We refer to $\rho_{i,n}$ (resp., $\zeta_{i,n}$) as seller i's price (resp., marginal cost) in period n. Consistently with our assumptions and parametrization, we treat each seller i's pricing problem as a single-agent problem. We denote by $\mathcal{G}_i^\infty$ seller i's empirical dynamic environment and by $\mathcal{G}_i(\lambda_i)$ seller i's empirical stage game.

ANR Pricing. On the marketplace, sellers face a trade-off: lower prices increase the probability of selling a device but lead to lower margins. To learn what prices are best for given marginal costs, each seller i faces a multi-armed bandit problem. Setting new prices allows the seller to discover the payoff of a different course of action (exploration) but at the cost of setting sub-optimal prices based on the available information (exploitation). ANR algorithms are specifically developed to address this exploration-exploitation trade-off. Hence, we assume that each seller i adopts some ANR algorithm to set prices.

5.3 Convergence to Set of BCCEs and Confidence Regions

Bayes Coarse Correlated Equilibrium. A BCCE of the empirical stage game $\mathcal{G}_i(\lambda_i)$ is a probability distribution over prices, marginal costs, and states satisfying certain restrictions.

Definition 11. *The probability distribution $\nu_i \in \Delta(A \times T \times \Theta)$ is a Bayes coarse correlated equilibrium of $\mathcal{G}(\lambda_i)$ if there exists some probability distribution $\tilde{\nu}_i \in \Delta(A \times T)$ such that, for all $\rho_i \in A$, $\zeta_i \in T$, and $\overline{\rho}_{-i} \in \Theta$, we have $\nu_i(\rho_i, \zeta_i, \overline{\rho}_{-i}) = \tilde{\nu}_i(\rho_i, \zeta_i)\psi(\overline{\rho}_{-i})$, and:*

[15] A discrete set of actions is routine in algorithmic environments.

1. *$\tilde{\nu}_i$ is consistent for $\mathcal{G}(\lambda_i)$; that is, for all $\zeta_i \in T$, we have $\sum_{\rho_i} \tilde{\nu}_i(\rho_i, \zeta_i) = \psi_i(\zeta_i; \lambda_{i,0})$.*
2. *$\tilde{\nu}_i$ is coarsely obedient for $\mathcal{G}(\lambda_i)$; that is, for all $\zeta_i \in T$, we have*

$$\sum_{\rho_i} \left[\sum_{\overline{\rho}_{-i}} \left[s_i(\rho_i', \overline{\rho}_{-i})(\rho_i' - \zeta_i) - s_i(\rho_i, \overline{\rho}_{-i})(\rho_i - \zeta_i) \right] \psi(\overline{\rho}_{-i}) \right] \tilde{\nu}_i(\rho_i, \zeta_i) \leq 0 \quad \textit{for all } \rho_i' \in A.$$

We denote by $E_i(\lambda_i)$ the set of BCCEs of $\mathcal{G}_i(\lambda_i)$.

We can write ν_i as a product distribution $\nu_i = \tilde{\nu}_i \times \psi$ for some $\tilde{\nu}_i \in \Delta(A \times T)$ because of our assumption that ζ_i and $\overline{\rho}_{-i}$ are independent, and so are ρ_i and $\overline{\rho}_{-i}$. Otherwise, the definition of a BCCE of $\mathcal{G}(\lambda_i)$ is standard and mimics Definition 3 in our general model.

Convergence under ANR Pricing. Since we assume that seller i adopts some ANR algorithm to set prices, the sequence of prices, marginal costs, and states $(\rho_{i,n}, \zeta_{i,n}, \overline{\rho}_{-i,n})_{n=1}^{\infty}$ from $\mathcal{G}_i^{\infty}$ has ANR. Hence, the equivalent of Theorem 1 in the context of our empirical model holds. We summarize the relevant implications for estimation with the following proposition.

Proposition 1. *Suppose the sequence of prices, marginal costs, and states $(\rho_{i,n}, \zeta_{i,n}, \overline{\rho}_{-i,n})_{n=1}^{\infty}$ from $\mathcal{G}_i^{\infty}$ has ANR. Then, the sequence of empirical distributions of prices, marginal costs, and states $(Z_{i,N})_{N=1}^{\infty}$, where $Z_{i,N}(\rho_i, \zeta_i, \overline{\rho}_{-i}) \coloneqq \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}_{\{\rho_i\}}(\rho_{i,n})\mathbb{1}_{\{\zeta_i\}}(\zeta_{i,n})\mathbb{1}_{\{\overline{\rho}_{-i}\}}(\overline{\rho}_{-i,n})$ for all $N \in \mathbb{N}$ and $(\rho_i, \zeta_i, \overline{\rho}_{-i}) \in A \times T \times \Theta$, converges almost surely to $E_i(\lambda_{i,0})$.*

Estimation, Confidence Region, and Coverage. Under our assumptions, for each seller i and period n, the payoff associated with each price is drawn independently from a fixed distribution that depends only on the price chosen. Such distribution is the same across periods so that the draws are i.i.d. Moreover, we assume that each seller i can observe the payoffs for the selected price and construct counterfactual payoffs for all other prices. Hence, each seller i faces a stochastic i.i.d. bandit environment with full feedback.[16] We further assume that sellers adopt an algorithm that optimizes the convergence rate of the expected regret to 0 and use these assumptions (see, e.g., Section 1.2 in Faure, Gaillard, Gaujal, and Perchet, 2015) to construct a uniformly valid confidence region.

For our purpose, we must introduce some notation. The identified set, denoted by Λ_i^*, is defined following Definition 7, accounting for the specifics of our empirical model. Let $\underline{\varepsilon} \coloneqq (\varepsilon(i, \zeta_i))_{\zeta_i \in T}$, where $\varepsilon(i, \zeta_i) \geq 0$ for all $\zeta_i \in T$. We denote by $Q_i(\underline{\varepsilon}; \lambda_i)$ the set of $\underline{\varepsilon}$-BCCE predictions of $\mathcal{G}_i(\lambda_i)$. Given the observed empirical distribution of N actions of seller i, denoted by $q_{i,N}$, the $\underline{\varepsilon}$-BCCE confidence region is $\widehat{\Lambda}_{i,N}(\underline{\varepsilon}) \coloneqq \{\lambda_i \in \Lambda : q_{i,N} \in Q_i(\underline{\varepsilon}; \lambda_i)\}$.

[16] In Appendix F.7, we relax the assumptions that $\overline{\rho}_{-i,n}$ are i.i.d., that $\zeta_{i,n}$ and $\overline{\rho}_{-i,n}$ are independent, and of full feedback. We show that our estimation results are very similar under alternative assumptions.

The following proposition, which is the equivalent of Theorem 3 for our empirical model, establishes that, by appropriately specifying $\varepsilon(i,\zeta_i)$ for all $\zeta_i \in T$, the set $\widehat{\Lambda}_{i,N}(\underline{\varepsilon})$ becomes a uniformly valid confidence region for all $\lambda_i \in \Lambda_i^*$ at any desired confidence level. To establish the result, we need a few final pieces of notation. Consider seller i with marginal cost ζ_i. The expected payoff from setting price ρ_i is $\mu(i,\rho_i,\zeta_i) := \sum_{\overline{\rho}_{-i}} s_i(\rho_i,\overline{\rho}_{-i})(\rho_i-\zeta_i)\psi(\overline{\rho}_{-i})$; the expected payoff from setting the best price is $\mu^*(i,\rho_i) := \max_{\rho_i\in T}\mu(i,\rho_i,\zeta_i)$; the gap of price ρ_i is $\Phi(i,\rho_i,\zeta_i) := \mu^*(i,\rho_i) - \mu(i,\rho_i,\zeta_i)$. Finally, define

$$K(i,\zeta_i) := \max_{\rho_i,\overline{\rho}_{-i},\rho_i',\overline{\rho}_{-i}'} \left| s_i(\rho_i,\overline{\rho}_{-i})(\rho_i-\zeta_i) - s_i(\rho_i',\overline{\rho}_{-i}')(\rho_i'-\zeta_i)\right|,$$

and

$$\Phi(i,\zeta_i) := \sum_{\substack{\rho_i\in A:\\ \Phi(i,\rho_i,\zeta_i)>0}} \frac{K(i,\zeta_i)}{\Phi(i,\rho_i,\zeta_i)}.$$

The following result holds.

Proposition 2. *Suppose the assumptions above on the empirical model hold. Fix $\alpha \in (0,1)$ and $N \in \mathbb{N}$. For all $\zeta_i \in T$, let*

$$\varepsilon_{\alpha,N}(i,\zeta_i) := \frac{\Phi(i,\zeta_i)}{\left[1-(1-\alpha)^{\frac{1}{|T|}}\right]N}.$$

Then, $\widehat{\Lambda}_{i,N}(\underline{\varepsilon}_{\alpha,N}) := \left\{\lambda_i \in \Lambda : q_{i,N} \in Q_i(\underline{\varepsilon}_{\alpha,N};\lambda_i)\right\}$ is a uniformly valid confidence region for all $\lambda_i \in \Lambda_i^*$; that is,

$$\inf_{\lambda_i\in\Lambda_i^*} \mathbb{P}\Big(\lambda_i \in \widehat{\Lambda}_{i,N}(\underline{\varepsilon}_{\alpha,N})\Big) \geq 1-\alpha.$$

We use Proposition 2 to obtain the estimates we present in the next sections.

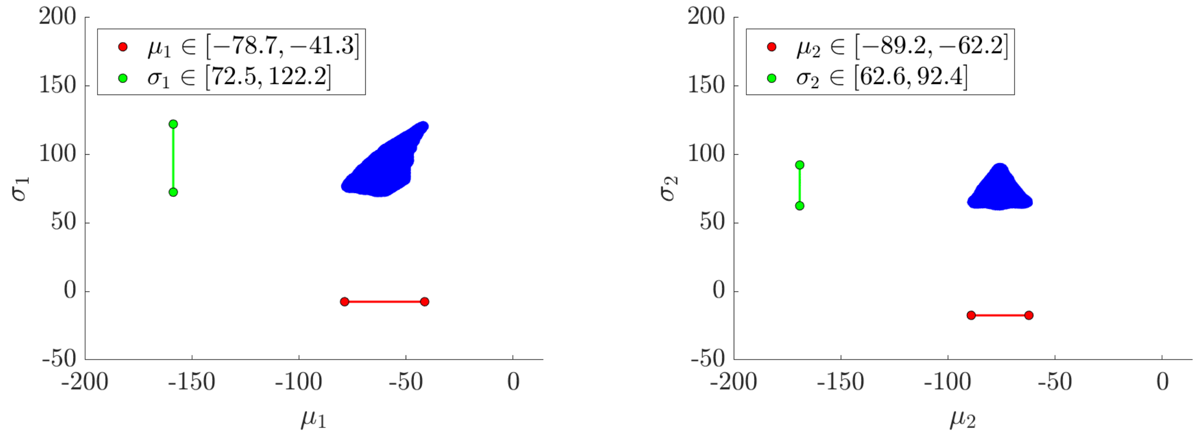
5.4 Estimation Results

Relying on Proposition 2, we compute for each seller i the confidence region $\widehat{\Lambda}_{i,N}(\underline{\varepsilon}_{\alpha,N})$ corresponding to $\alpha = 0.05$. The computation of the confidence regions specializes the results in Section 4.4 to the context of our empirical model. Details are in Appendix E.3.

In panels (a) and (b) of Figure 7, we represent the confidence regions for the top-2 Swappa sellers in our data. For each seller i, the figure plots (in blue) the confidence region $\widehat{\Lambda}_{i,N}(\underline{\varepsilon}_{\alpha,N})$ in the (μ_i,σ_i)-space, the parameters characterizing the seller-specific distribution of marginal costs. We also represent (in red and green) the projections of the confidence regions for each of the two parameters.

The confidence regions vary in shape and size across sellers, reflecting different business strategies across sellers. Parameter μ_i ranges between -89 and -41; parameter σ_i ranges

Figure 7: Confidence Regions.



(a) Seller 1

(b) Seller 2

We plot in blue confidence regions in the (μ_i, σ_i)-space; we report in red and green projections for μ_i and σ_i. Each panel corresponds to one of the top 2 Swappa sellers. Computational details are in Appendix E.3.

between 62 and 122. The parameters' signs conform with intuition; in particular, the negative μ_i reflects that the distribution of cost residuals takes on mostly negative values, as marginal costs for sellers are typically below the reference prices of devices.

To better understand the implications of the confidence regions, we simulate costs from the estimated distribution $\psi_i(\cdot; \lambda_i)$ across values of λ_i in the confidence region. We report marginal cost statistics in Table 1. Columns 1 and 2 report statistics of marginal cost residuals, translating the parameters in the confidence regions in realizations of $\zeta_{i,n}$. In line with the nature of the truncated, discretized Normal distribution we use, mean marginal cost residuals are close to the estimated μ_i and in the range of $-\$90$ to $-\$56$. These dollar amounts represent the deviations of sellers' marginal costs from reference prices. For comparison, average reference prices, depending on the seller, range from \$527 to \$596 (column 6). Therefore, mean marginal costs range between \$436 and \$540 for Sellers 1 and 2. The variability in marginal cost residuals adds little variation to marginal cost (column 4), when compared to the variability in reference prices (column 6). Sellers bring to Swappa a wide variety of device models, ranging from more expensive recent models to older, cheaper devices. The range of marginal costs implied by our estimates seems sensible and in line with the economics of reselling used electronics: procuring and refurbishing most used devices costs around $\$50 - \$100$ less than reference sale prices in centralized marketplaces.

Recovering marginal costs enables a range of measurement exercises. First, we can recover average markups, i.e., the average difference between prices and marginal costs. This object is crucial to evaluating the competitiveness of markets. How tough is the competition

Table 1: Estimated Distributions of Marginal Cost Residuals and Marginal Costs.

	Marginal Cost Residuals ($)		Marginal Cost ($)		Reference Price ($)	
	Mean	SD	Mean	SD	Mean	SD
	(1)	(2)	(3)	(4)	(5)	(6)
Seller 1	$[-83.5, -56.2]$	$[69.0, 98.0]$	$[512.4, 539.7]$	$[196.9, 218.6]$	595.9	191.7
Seller 2	$[-90.8, -62.3]$	$[60.2, 86.5]$	$[436.2, 464.7]$	$[254.3, 271.1]$	527.0	253.0

We report statistics of marginal cost residuals (columns 1-2), marginal costs (columns 3-4), and reference prices (columns 5-6). Each of the top-2 Swappa sellers corresponds to a row. Marginal cost residual statistics are obtained by simulating 2,000 draws of $\zeta_{i,n}$ from $\psi_i(\cdot;\lambda_i)$ for all $\lambda_i \in \widehat{\Lambda}_{i,N}(\underline{\varepsilon}_{\alpha,N})$. We then take the minimum and maximum of the statistics across $\lambda_i \in \widehat{\Lambda}_{i,N}(\underline{\varepsilon}_{\alpha,N})$ to generate the interval in the table. Mean marginal costs (column 3) are obtained by summing average reference prices in the data (column 4) with average marginal cost residuals (column 1). To compute the standard deviation of marginal cost (column 4), we draw reference prices from their empirical distribution.

among sellers on Swappa? Are sellers on a decentralized platform charging lower markups than centralized sellers? We turn to these questions in the next section.

While we do not pursue those here, using the recovered marginal cost primitive would also permit us to simulate pricing on Swappa under a range of counterfactual platform designs and models of seller behavior. For instance, we could obtain outcomes given the adoption by sellers of a specific class of AIs (not necessarily within the ANR class). This policy experiment would mimic the implementation of "suggested pricing" by the platform, whereby the marketplace makes algorithmic pricing tools available to sellers.

5.5 Quantification of Markups

Measuring markups is the most direct way of assessing firms' market power. For data generated in environments such as digital platforms, where sellers often use AIs to help them set prices, our method provides a way of recovering markups under an assumption (ANR) that is more suitable than standard oligopoly conduct assumptions typically adopted under the "demand approach" to the estimation of markups.

We use the marginal cost estimates obtained with our method to evaluate markups for sellers on Swappa. We define the average markup for seller i as $\Delta_i(\lambda_i) := \mathbb{E}[p_{i,n}] - \mathbb{E}_{\psi_i(\cdot;\lambda_i)}[t_{i,n}]$, and construct the intervals

$$\left[\min_{\lambda_i \in \widehat{\Lambda}_{i,N}(\underline{\varepsilon}_{\alpha,N})} \Delta_i(\lambda_i), \max_{\lambda_i \in \widehat{\Lambda}_{i,N}(\underline{\varepsilon}_{\alpha,N})} \Delta_i(\lambda_i)\right].$$

We compare markups for Swappa sellers to those on Gazelle, a centralized buyback and resale platform for used smartphones.[17] To assess markups on Gazelle, we use that we can observe device-model specific sale ("ask") prices and the trade-in ("bid") prices offered by the platform. The bid prices do not coincide with Gazelle's marginal cost, as the firm needs to pay for shipping, inspecting, and refurbishing the devices it receives. Since we do not have direct access to cost or financial information from Gazelle, we use industry sources to estimate these costs between 10% and 30% of the bid price. We compute a range for Gazelle markups by taking the difference between ask and bid prices adjusted by the cost multipliers 1.1 and 1.3. Details on Gazelle data are in Appendix F.5.

Table 2: Markups on Swappa and Gazelle

		Device		
	All Devices	iPhone 13 128GB Mint	iPhone 12 64GB Good	iPhone 11 64GB Good
	(1)	(2)	(3)	(4)
Mean Swappa Price ($)	505.5	553.1	343.1	276.4
Mean Swappa Markups ($)				
Seller 1 Markup	[20.5, 44.5]	[28.7, 52.7]	[51.5, 75.5]	[52.4, 76.4]
Seller 2 Markup	[24.7, 50.9]	[42.4, 68.6]	[37.2, 63.4]	[32.2, 58.4]
Gazelle Markup ($)	·	[81.8, 135.7]	[84.4, 114.3]	[58.3, 82.6]

We report statistics of prices and markups across all devices (column (1)) and for the three best-selling devices for three popular iPhone models (columns (2)–(4)) in our data. Swappa prices are averages for sold devices. Seller-specific mean Swappa markups are obtained by subtracting the average marginal costs from prices. Gazelle markups are obtained by subtracting the bid prices times a cost multiplier from Gazelle ask prices. We consider multipliers of 1.1 and 1.3 to give a conservative interval.

Table 2 summarizes our markup quantification exercise. Average markups for Swappa sellers are in the range of \$21–\$51. These markups vary across sellers and device models: for instance, Seller 1 can charge average markups up to \$75 for popular models of iPhone 11 and 12 but can only charge average markups of \$29–\$53 on a newer iPhone 13. Seller 2, instead, sets average markups of \$42–\$68 on the same device while earning lower markups on older devices. Overall, markups on Gazelle are higher than those charged by Swappa sellers, particularly for newer devices. For the model of iPhone 13 we examine, Gazelle markups range between \$82–\$136, above the interval for Swappa sellers 1 and 2. Similarly, Gazelle markups for the older iPhones 11 and 12 are above the average markups charged by the two largest Swappa sellers.

In sum, margins for Swappa sellers are modest. This finding is not surprising, given the

[17] See https://buy.gazelle.com/. Gazelle is similar to Decluttr, from which we obtained reference prices. Decluttr stopped accepting trade-ins during our sample period, thus preventing the evaluation of its markups.

low barriers to entry for retailers on the platform. Our measurement exercise supports the view that Gazelle, a centralized platform that operates on a larger scale than independent Swappa sellers, can sustain higher markups, especially for newer devices. Our results, however, do not speak to the source of these differences: perhaps consumers are less elastic when buying from Gazelle, or possibly the ability to purchase trade-ins directly from its website gives a marginal cost advantage to Gazelle. Overall, however, this application showcases our methods' potential to deliver credible estimates of economic primitives in environments where AIs may be setting prices.

6 Conclusion

We develop a novel approach to estimate a game's primitives in online environments where AI supports or replaces human decisions. Given the theoretical and practical relevance of regret-minimizing algorithms in online learning, we assume that algorithms in this broad class generate the data. Thus, our model departs from standard equilibrium assumptions.

We show that under asymptotic no regret, the empirical distribution of actions converges to the set of Bayes coarse correlated equilibrium predictions of the underlying stage game. This connection lets us derive econometric properties directly from the theoretical convergence results. First, we define a meaningful notion of an identified set in this context and propose an estimator based on Bayes coarse correlated equilibrium restrictions. Second, we develop an inferential procedure that builds on convergence rates of regret-minimizing algorithms, allowing us to construct confidence regions with desired coverage properties.

Our empirical application to the Swappa marketplace demonstrates the practical value of this approach. We recover the distribution of sellers' marginal costs, quantify markups, and compare them to those on competing centralized platforms. This analysis provides insights into the competitiveness of decentralized digital marketplaces and showcases how our method enables various measurement and counterfactual exercises. This paper is a first step in studying interactions among AIs with a structural econometric approach, and we anticipate a wide range of applications to policy and market design questions in AI-driven environments.

References

Abada, I. and X. Lambin (2023): "Artificial Intelligence: Can Seemingly Collusive Outcomes be Avoided?" *Management Science*, 69, 5042–5065.

Aguirregabiria, V. and J. Jeon (2020): "Firms' Beliefs and Learning: Models, Identification, and Empirical Evidence," *Review of Industrial Organization*, 56, 203–235.

Aguirregabiria, V. and A. Magesan (2020): "Identification and Estimation of Dynamic Discrete Games When Players' Beliefs are Not in Equilibrium," *The Review of Economic Studies*, 87, 582–625.

Aouad, A. and A. V. den Boer (2021): "Algorithmic Collusion in Assortment Games," *Working Paper*.

Arora, S., E. Hazan, and S. Kale (2012): "The Multiplicative Weights Update Method: A Meta-Algorithm and Applications," *Theory of Computing*, 8, 121–164.

Asker, J., C. Fershtman, and A. Pakes (2024): "The Impact of Artificial Intelligence Design on Pricing," *Journal of Economics & Management Strategy*, 33, 276–304.

Assad, S., R. Clark, D. Ershov, and L. Xu (2024): "Algorithmic Pricing and Competition: Empirical Evidence from the German Retail Gasoline Market," *Journal of Political Economy*, 132, 723–771.

Auer, P., N. Cesa-Bianchi, Y. Freund, and R. E. Schapire (2002): "The Nonstochastic Multiarmed Bandit Problem," *SIAM Journal on Computing*, 32, 48–77.

Aumann, R. J. (1974): "Subjectivity and Correlation in Randomized Strategies," *Journal of Mathematical Economics*, 1, 67–96.

——— (1987): "Correlated Equilibrium as an Expression of Bayesian Rationality," *Econometrica*, 55, 1–18.

Banchio, M. and G. Mantegazza (2024): "Artificial Intelligence and Spontaneous Collusion," *Working Paper*.

Banchio, M. and A. Skrzypacz (2022): "Market Design for AI Algorithms," *ACM SIGecom Exchanges*, 20, 62–68.

Bergemann, D. and S. Morris (2013): "Robust Predictions in Games with Incomplete Information," *Econometrica*, 81, 1251–1308.

——— (2016): "Bayes Correlated Equilibrium and the Comparison of Information Structures in Games," *Theoretical Economics*, 11, 487–522.

Blackwell, D. (1956a): "An Analog of the Minmax Theorem for Vector Payoffs," *Pacific Journal of Mathematics*, 6, 1–8.

——— (1956b): "Controlled Random Walks," *Proceedings of the International Congress of Mathematicians*, 3, 336–338.

Blum, A. and Y. Mansour (2007): "From External to Internal Regret." *Journal of Machine Learning Research*, 8.

Brooks, B., S. Du, and L. Zhang (2024): "Participation-Adaptive Pricing," *Working Paper.*

Brown, Z. Y. and A. MacKay (2023): "Competition in Pricing Algorithms," *American Economic Journal: Microeconomics*, 15, 109–256.

Bubeck, S. and N. Cesa-Bianchi (2012): "Regret Analysis of Stochastic and Nonstochastic Multi-Armed Bandit Problems," *Foundations and Trends® in Machine Learning*, 5, 1–122.

Calvano, E., G. Calzolari, V. Denicolò, and S. Pastorello (2020): "Artificial Intelligence, Algorithmic Pricing, and Collusion," *The American Economic Review*, 110, 3267–97.

Caragiannis, I., C. Kaklamanis, P. Kanellopoulos, M. Kyropoulou, B. Lucier, R. Paes Leme, and E. Tardos (2015): "Bounding the Inefficieny of Outcomes in Generalized Second Price Auctions," *Journal of Economic Theory*, 157, 343–388.

Cesa-Bianchi, N. and G. Lugosi (2006): *Prediction, Learning, and Games*, Cambridge University Press.

Chassang, S. and J. Ortner (2023): "Regulating Collusion," *Annual Review of Economics*, 15, 177–204.

Chen, L., A. Mislove, and C. Wilson (2016): "An Empirical Analysis of Algorithmic Pricing on Amazon Marketplace," in *Proceedings of the 25th International Conference on World Wide Web*, 1339–1349.

Chen, M., S. Nabi, and M. Siniscalchi (2023): "Advancing Ad Auction Realism: Practical Insights & Modeling Implications," *Working Paper.*

Doraszelski, U., G. Lewis, and A. Pakes (2018): "Just Starting Out: Learning and Equilibrium in a New Market," *The American Economic Review*, 108, 565–615.

Epstein, L., H. Kaido, and K. Seo (2016): "Robust Confidence Regions for Incomplete Models," *Econometrica*, 84, 1799–1838.

Farina, G., J. Grand-Clément, C. Kroer, C.-W. Lee, and H. Luo (2024): "Regret Matching$^+$: (In)Stability and Fast Convergence in Games," *Advances in Neural Information Processing Systems*, 36.

Faure, M., P. Gaillard, B. Gaujal, and V. Perchet (2015): "Online Learning and Game Theory. A Quick Overview with Recent Results and Applications," *ESAIM: Proceedings and Surveys*, 51, 246–271.

Foster, D. P. and R. V. Vohra (1997): "Calibrated Learning and Correlated Equilibrium," *Games and Economic Behavior*, 21, 40–55.

——— (1999): "Regret in the On-Line Decision Problem," *Games and Economic Behavior*, 29, 7–35.

Freund, Y. and R. E. Schapire (1997): "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," *Journal of Computer and System Sciences*, 55, 119–139.

Gradwohl, R. and E. Kalai (2021): "Large Games: Robustness and Stability," *Annual Review of Economics*, 13.

Gualdani, C. and S. Sinha (2024): "Identification and Inference in Discrete Choice Models with Imperfect Information," *Journal of Econometrics*, 244, 105854.

Haile, P. A. and E. Tamer (2003): "Inference with an Incomplete Model of English Auctions," *Journal of Political Economy*, 111, 1–51.

Hannan, J. (1957): "Approximation to Bayes Risk in Repeated Play," in *Contributions to the Theory of Games, Vol. 3*, ed. by M. Dresher, A. Tucker, and P. Wolfe, Princeton University Press.

Hansen, K. T., K. Misra, and M. M. Pai (2021): "Frontiers: Algorithmic Collusion: Supra-Competitive Prices via Independent Algorithms," *Marketing Science*, 40, 1–12.

Hart, S. and A. Mas-Colell (2000): "A Simple Adaptive Procedure Leading to Correlated Equilibrium," *Econometrica*, 68, 1127–1150.

——— (2001): "A Reinforcement Procedure Leading to Correlated Equilibrium," in *Economic Essays: A Festschrift for Werner Hildenbrand*, ed. by G. Debreu, W. Neuefeind, and W. Trockel, Springer.

——— (2013): *Simple Adaptive Strategies: From Regret-Matching to Uncoupled Dynamics*, World Scientific.

Hartline, J., V. Syrgkanis, and E. Tardos (2015): "No-Regret Learning in Bayesian Games," *NIPS 2015: Advances in Neural Information Processing Systems*, 3061–3069.

Hartline, J. D., S. Long, and C. Zhang (2024): "Regulation of algorithmic collusion," in *Proceedings of the Symposium on Computer Science and Law*, 98–108.

Hazan, E. (2022): *Introduction to Online Convex Optimization*, MIT Press.

Immorlica, N., B. Lucier, and A. Slivkins (2024): "Generative AI as Economic Agents," *ACM SIGecom Exchanges*, 22, 93–109.

Jensen, M. K. (2018): "Aggregative Games," in *Handbook of Game Theory and Industrial Organization, Volume I*, Edward Elgar Publishing.

Kalai, A. and S. Vempala (2005): "Efficient Algorithms for Online Decision Problems," *Journal of Computer and System Sciences*, 71, 291–307.

Lamba, R. and S. Zhuk (2024): "Pricing with Algorithms," *Working Paper.*

Lattimore, T. and C. Szepesvári (2020): *Bandit Algorithms*, Cambridge University Press.

Littlestone, N. and M. K. Warmuth (1994): "The Weighted Majority Algorithm," *Information and Computation*, 108, 212–261.

Magnolfi, L. and C. Roncoroni (2023): "Estimation of Discrete Games with Weak Assumptions on Information," *The Review of Economic Studies*, 90, 2006–2041.

Manski, C. F. (2003): *Partial Identification of Probability Distributions*, Springer Science & Business Media.

Milgrom, P. R. and S. Tadelis (2018): "How Artificial Intelligence and Machine Learning Can Impact Market Design," in *The Economics of Artificial Intelligence: An Agenda*, University of Chicago Press, 567–585.

Molinari, F. (2020): "Microeconometrics with Partial Identification," *Handbook of Econometrics*, 7, 355–486.

Moulin, H. and J.-P. Vial (1978): "Strategically Zero-Sum Games: The Class of Games whose Completely Mixed Equilibria Cannot Be Improved Upon," *Internation Journal of Game Theory*, 7, 201–221.

Musolff, L. (2024): "Algorithmic Pricing Facilitates Tacit Collusion: Evidence from E-Commerce," *Working Paper*.

Nekipelov, D., V. Syrgkanis, and E. Tardos (2015): "Econometrics for Learning Agents," in *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, 1–18.

Nemirovsky, A. S. and D. B. Yudin (1983): *Problem Complexity and Method Efficiency in Optimization*, Wiley-Interscience.

Nisan, N., T. Roughgarden, E. Tardos, and V. V. Vazirani (2007): *Algorithmic Game Theory*, Cambridge University Press.

Noti, G. and V. Syrgkanis (2021): "Bid Prediction in Repeated Auctions with Learning," in *Proceedings of the Web Conference 2021*, 3953–3964.

Roughgarden, T. (2016): *Twenty Lectures on Algorithmic Game Theory*, Cambridge University Press.

Salcedo, B. (2015): "Pricing Algorithms and Tacit Collusion," *Working Paper*.

Shalev-Shwartz, S. (2012): "Online Learning and Online Convex Optimization," *Foundations and Trends® in Machine Learning*, 4, 107–194.

Slivkins, A. (2019): "Introduction to Multi-Armed Bandits," *Foundations and Trends® in Machine Learning*, 12, 1–286.

Syrgkanis, V., E. Tamer, and J. Ziani (2021): "Inference on Auctions with Weak Assumptions on Information," *Working Paper*.

Tamer, E. (2003): "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *The Review of Economic Studies*, 70, 147–165.

Weintraub, G. Y., C. L. Benkard, and B. Van Roy (2008): "Markov Perfect Industry Dynamics with Many Firms," *Econometrica*, 76, 1375–1411.

Zhang, L. (2024): "Revenue Guarantee across Bayes Coarse Correlated Equilibria," *Working Paper*.

Appendix

A Proofs for Sections 3 and 4

A.1 Proof of Theorem 1

[$\Longrightarrow$] Suppose the sequence $(a_n, t_n, \theta_n)_{n=1}^{\infty}$ from $\mathcal{G}^{\infty}$ has ANR. Consider any subsequence $(Z_{N_l})_{l=1}^{\infty}$ of $(Z_N)_{N=1}^{\infty}$ that converges to some $\nu \in \Delta(A \times T \times \Theta)$. We need to show that ν is almost surely consistent and coarsely obedient for $\mathcal{G}(\lambda_0)$.

Consistency. Pick any $(t, \theta) \in T \times \Theta$. Note that

$$\begin{aligned}
\sum_a \nu(a,t,\theta) &= \sum_a \lim_{l\to\infty} Z_{N_l}(a,t,\theta) \\
&= \lim_{l\to\infty} \sum_a Z_{N_l}(a,t,\theta) \\
&= \lim_{l\to\infty} \left[\frac{\sum_a Z_{N_l}(a,t,\theta)}{\sum_{a,t} Z_{N_l}(a,t,\theta)} \sum_{a,t} Z_{N_l}(a,t,\theta) \right] \qquad (3) \\
&= \lim_{l\to\infty} \frac{\sum_a Z_{N_l}(a,t,\theta)}{\sum_{a,t} Z_{N_l}(a,t,\theta)} \lim_{l\to\infty} \sum_{a,t} Z_{N_l}(a,t,\theta) \\
&= \lim_{l\to\infty} \frac{\sum_{n=1}^{N_l} \mathbb{1}_{\{t\}}(t_n)\mathbb{1}_{\{\theta\}}(\theta_n)}{\sum_{n=1}^{N_l} \mathbb{1}_{\{\theta\}}(\theta_n)} \lim_{l\to\infty} \frac{\sum_{n=1}^{N_l} \mathbb{1}_{\{\theta\}}(\theta_n)}{N_l}.
\end{aligned}$$

The ratio

$$\frac{\sum_{n=1}^{N_l} \mathbb{1}_{\{t\}}(t_n)\mathbb{1}_{\{\theta\}}(\theta_n)}{\sum_{n=1}^{N_l} \mathbb{1}_{\{\theta\}}(\theta_n)} \qquad (4)$$

is the empirical frequency of signal profile t at periods when the state is θ. Since signal profiles t_n are drawn from $\pi(\cdot \,|\, \theta_n; \lambda_0)$, the ratio in (4) is the empirical frequency of $\sum_{n=1}^{N_l} \mathbb{1}_{\{\theta\}}(\theta_n)$ conditionally independent draws from $\pi(\cdot \,|\, \theta; \lambda_0)$. Moreover, since $\psi(\cdot; \lambda_0) \in \Delta_{++}(\Theta)$, we have $\lim_{l\to\infty} \sum_{n=1}^{N_l} \mathbb{1}_{\{\theta\}}(\theta_n) = \infty$ almost surely. Thus, by the strong law of large numbers,

$$\lim_{l\to\infty} \frac{\sum_{n=1}^{N_l} \mathbb{1}_{\{t\}}(t_n)\mathbb{1}_{\{\theta\}}(\theta_n)}{\sum_{n=1}^{N_l} \mathbb{1}_{\{\theta\}}(\theta_n)} = \pi(t \,|\, \theta; \lambda_0) \quad \text{a.s.} \qquad (5)$$

Moreover, since states are i.i.d. draws from $\psi(\cdot; \lambda_0)$, again by the strong law of large numbers,

$$\lim_{l\to\infty} \frac{\sum_{n=1}^{N_l} \mathbb{1}_{\{\theta\}}(\theta_n)}{N_l} = \psi(\theta; \lambda_0) \quad \text{a.s.} \qquad (6)$$

From equations (3), (5), and (6), we have

$$\sum_a \nu(a,t,\theta) = \pi(t \,|\, \theta; \lambda_0)\psi(\theta; \lambda_0) \quad \text{a.s.} \qquad (7)$$

We conclude from equation (7) that ν is almost surely consistent for $\mathcal{G}(\lambda_0)$.

Coarse obedience. Pick any $i \in \mathcal{I}$ and $t_i \in T_i$. For all $a_i' \in A_i$, let

$$V_N(i, t_i, a_i') \coloneqq \frac{1}{N} \sum_{n=1}^{N} u_i\big((a_i', a_{-i,n}), \theta_n; \lambda_0\big) \mathbb{1}_{\{t_i\}}(t_{i,n})$$

be the payoff of agent i with signal t_i for not having always played action a_i' up to period N. Note that

$$\begin{aligned} &V_N(i, t_i, a_i') - U_N(i, t_i) \\ &\quad = \frac{1}{N} \sum_{n=1}^{N} \big[u_i\big((a_i', a_{-i,n}), \theta_n; \lambda_0\big) - u_i\big((a_{i,n}, a_{-i,n}), \theta_n; \lambda_0\big)\big] \mathbb{1}_{\{t_i\}}(t_{i,n}) \\ &\quad = \frac{1}{N} \sum_{n=1}^{N} \sum_{a, t_{-i}, \theta} \big[u_i\big((a_i', a_{-i}), \theta; \lambda_0\big) - u_i(a, \theta; \lambda_0)\big] \mathbb{1}_{\{a\}}(a_n) \mathbb{1}_{\{t_i\}}(t_{i,n}) \mathbb{1}_{\{t_{-i}\}}(t_{-i,n}) \mathbb{1}_{\{\theta\}}(\theta_n) \\ &\quad = \sum_{a, t_{-i}, \theta} \big[u_i\big((a_i', a_{-i}), \theta; \lambda_0\big) - u_i(a, \theta; \lambda_0)\big] Z_N\big(a, (t_i, t_{-i}), \theta\big). \end{aligned} \tag{8}$$

Since $(a_n, t_n, \theta_n)_{n=1}^{\infty}$ has ANR, for all $a_i' \in A_i$, we have

$$\limsup_{N \to \infty} \big[V_N(i, t_i, a_i') - U_N(i, t_i)\big] \le 0 \quad \text{a.s.} \tag{9}$$

Then, by equation (8) and inequality (9), for all $a_i' \in A_i$, we have

$$\limsup_{N \to \infty} \sum_{a, t_{-i}, \theta} \big[u_i\big((a_i', a_{-i}), \theta; \lambda_0\big) - u_i(a, \theta; \lambda_0)\big] Z_N\big(a, (t_i, t_{-i}), \theta\big) \le 0 \quad \text{a.s.} \tag{10}$$

Moreover, on the subsequence $(Z_{N_l})_{l=1}^{\infty}$, for all $a_i' \in A_i$, we have

$$\begin{aligned} &\lim_{l \to \infty} \sum_{a, t_{-i}, \theta} \big[u_i\big((a_i', a_{-i}), \theta; \lambda_0\big) - u_i(a, \theta; \lambda_0)\big] Z_{N_l}\big(a, (t_i, t_{-i}), \theta\big) \\ &= \sum_{a, t_{-i}, \theta} \lim_{l \to \infty} \big[u_i\big((a_i', a_{-i}), \theta; \lambda_0\big) - u_i(a, \theta; \lambda_0)\big] Z_{N_l}\big(a, (t_i, t_{-i}), \theta\big) \\ &= \sum_{a, t_{-i}, \theta} \big[u_i\big((a_i', a_{-i}), \theta; \lambda_0\big) - u_i(a, \theta; \lambda_0)\big] \nu\big(a, (t_i, t_{-i}), \theta\big). \end{aligned} \tag{11}$$

By inequality (10) and equation (11), for all $a_i' \in A_i$, we have

$$\sum_{a, t_{-i}, \theta} \big[u_i\big((a_i', a_{-i}), \theta; \lambda_0\big) - u_i(a, \theta; \lambda_0)\big] \nu\big(a, (t_i, t_{-i}), \theta\big) \le 0 \quad \text{a.s.} \tag{12}$$

We conclude from equation (12) that ν is almost surely coarsely obedient for $\mathcal{G}(\lambda_0)$.

[$\Longleftarrow$] Now suppose $(Z_N)_{N=1}^{\infty}$ converges almost surely to $E(\lambda_0)$. Pick any $i \in \mathcal{I}$ and $t_i \in T_i$.

By coarse obedience, for all $a_i' \in A_i$, we have

$$\limsup_{N\to\infty} \sum_{a,t_{-i},\theta} \big[u_i\big((a_i', a_{-i}), \theta; \lambda_0\big) - u_i(a,\theta)\big] Z_N\big(a, (t_i, t_{-i}), \theta\big) \le 0 \quad \text{a.s.} \tag{13}$$

By equation (8) and inequality (13), we have $\limsup_{N\to\infty} \big[V_N(i,t_i,a_i') - U_N(i,t_i)\big] \le 0$ a.s. for all $a_i' \in A_i$, which implies $\limsup_{N\to\infty} R_N(i,t_i) \le 0$ a.s. The desired result follows. ■

A.2 Proof of Theorem 2

Pick any $\varepsilon > 0$ and $\lambda \in \Lambda^*$. Note that

$$\begin{aligned}
\lambda \in \Lambda^* &\Longrightarrow \mathbb{P}\Big(\lim_{N\to\infty} d(q_N, Q(\lambda)) = 0\Big) = 1\\
&\Longrightarrow \mathbb{P}\big(\text{for all } \tilde\varepsilon > 0,\ \text{there exists } N_{\tilde\varepsilon} \in \mathbb{N} \text{ s.th. } d(q_N, Q(\lambda)) < \tilde\varepsilon \text{ for all } N > N_{\tilde\varepsilon}\big) = 1\\
&\Longrightarrow \mathbb{P}\big(\text{for all } \tilde\varepsilon > 0,\ \text{there exists } N_{\tilde\varepsilon} \in \mathbb{N} \text{ s.th. } q_N \in Q(\lambda; \tilde\varepsilon) \text{ for all } N > N_{\tilde\varepsilon}\big) = 1\\
&\Longrightarrow \mathbb{P}\big(\text{there exists } N_{\varepsilon} \in \mathbb{N} \text{ s.th. } q_N \in Q(\varepsilon; \lambda) \text{ for all } N \ge N_{\varepsilon}\big) = 1\\
&\Longrightarrow \mathbb{P}\big(\text{there exists } N_{\varepsilon} \in \mathbb{N} \text{ s.th. } \lambda \in \widehat\Lambda_N(\varepsilon) \text{ for all } N \ge N_{\varepsilon}\big) = 1\\
&\Longrightarrow \mathbb{P}\left(\lambda \in \bigcup_{N=1}^{\infty} \bigcap_{N=K}^{\infty} \widehat\Lambda_N(\varepsilon)\right)\\
&\Longrightarrow \mathbb{P}\Big(\lambda \in \liminf_{N\to\infty} \lambda \in \widehat\Lambda_N(\varepsilon)\Big) = 1,
\end{aligned}$$

where: the first implication holds by the definition of identified set Λ^*; the second implication holds by the definition of the limit of a sequence of real numbers; the third implication holds by the definition of set of $\tilde\varepsilon$-BCCE predictions $Q(\lambda; \tilde\varepsilon)$; the fourth implication follows from the monotonicity of probability measures; the fifth implication holds by the definition of ε-BCCE estimator $\widehat\Lambda_N(\varepsilon)$; the last two implications hold by the definitions of set union, set intersection, and $\liminf$ of a sequence of sets. The desired result follows. ■

A.3 Proof of Theorem 3

For all $\lambda \in \Lambda$ and $\underline{\boldsymbol{\varepsilon}} \coloneqq (\varepsilon(i,t_i))_{i\in\mathcal{I}, t_i\in T_i}$, where $\varepsilon(i,t_i) \ge 0$ for all $i \in \mathcal{I}$ and $t_i \in T_i$, we have

$$\lambda \in \widehat\Lambda_N(\underline{\boldsymbol{\varepsilon}}) \Longleftrightarrow q_N \in Q(\underline{\boldsymbol{\varepsilon}}; \lambda) \Longleftrightarrow R_N(i,t_i) \le \varepsilon(i,t_i) \text{ for all } i \in \mathcal{I} \text{ and } t_i \in T_i, \tag{14}$$

where: the first equivalence holds by the definition of confidence region $\widehat\Lambda_N(\underline{\boldsymbol{\varepsilon}})$; the second equivalence holds by the definitions of the set of $\underline{\boldsymbol{\varepsilon}}$-BCCE predictions $Q(\underline{\boldsymbol{\varepsilon}}; \lambda)$ and regret $R_N(i,t_i)$, Theorem 1, and Corollary 1. For all $i \in \mathcal{I}$, $t_i \in T_i$, and $N \in \mathbb{N}$, let

$$N_{t_i} \coloneqq \sum_{n=1}^{N} \mathbb{1}_{\{t_i\}}(t_{i,n}) \tag{15}$$

be the number of times agent i observed signal t_i up to period N, and define

$$\overline{R}_N(i,t_i) \coloneqq N R_N(i,t_i). \tag{16}$$

Pick any $\lambda \in \Lambda^*$. Under assumptions (a)–(c) of the theorem, we have (see, e.g., Chapter 17 in Roughgarden (2016))

$$\mathbb{E}\big[\overline{R}_N(i,t_i) \mid N_{t_i}\big] \leq K(i,t_i;\lambda)\sqrt{\ln|A_i|}\sqrt{N}_{t_i}. \tag{17}$$

Next, note that

$$\begin{aligned}
\mathbb{E}\big[\overline{R}_N(i,t_i)\big] &= \mathbb{E}\big[\mathbb{E}\big[\overline{R}_{N_{t_i}}(i,t_i) \mid N_{t_i}\big]\big] \\
&\leq K(i,t_i;\lambda)\sqrt{\ln|A_i|}\,\mathbb{E}\big[\sqrt{N}_{t_i}\big] \\
&\leq K(i,t_i;\lambda)\sqrt{\ln|A_i|}\sqrt{\mathbb{E}[N_{t_i}]} \\
&= K(i,t_i;\lambda)\sqrt{\ln|A_i|}\sqrt{N\phi(t_i;\lambda)},
\end{aligned} \tag{18}$$

where: the first equality follows from the law of total expectation; the first inequality follows from inequality (17) and the linearity of expectation; the last inequality follows from Jensen's inequality; the last equality holds by the definition of N_{t_i} in (15) and the linearity of expectation, by observing that $(\mathbb{1}_{\{t_i\}}(t_{i,n}))_{n=1}^{\infty}$ is a sequence of i.i.d. random variables (because so is $(\theta_n)_{n=1}^{\infty}$) with $\mathbb{E}[\mathbb{1}_{\{t_i\}}(t_{i,n})] = \phi(t_i;\lambda)$.

By the definition of $\overline{R}_N(i,t_i)$ in (16), the chain of equalities and inequalities (18), and the linearity of expectation, we have

$$\mathbb{E}[R_N(i,t_i)] \leq \frac{K(i,t_i;\lambda)\sqrt{\ln|A_i|}\sqrt{\phi(t_i;\lambda)}}{\sqrt{N}}. \tag{19}$$

By inequality (19) and Markov's inequality, for all $\varepsilon > 0$, we have

$$\mathbb{P}(R_N(i,t_i) \leq \varepsilon) \geq 1 - \frac{K(i,t_i;\lambda)\sqrt{\ln|A_i|}\sqrt{\phi(t_i;\lambda)}}{\varepsilon\sqrt{N}}. \tag{20}$$

By inequality (20), for all $\overline{\alpha} \in (0,1)$,

$$\mathbb{P}(R_N(i,t_i) \leq \varepsilon(i,t_i)) \geq 1 - \overline{\alpha} \iff \varepsilon(i,t_i) = \frac{K(i,t_i;\lambda)\sqrt{\ln|A_i|}\sqrt{\phi(t_i;\lambda)}}{\overline{\alpha}\sqrt{N}}. \tag{21}$$

For the choice of $\varepsilon(i,t_i)$ on the right-hand side of equivalence (21), we have

$$\begin{aligned}
&\mathbb{P}\big(R_N(i,t_i) \leq \varepsilon(i,t_i) \text{ for all } i \in \mathcal{I} \text{ and } t_i \in T_i\big) \\
&\qquad = 1 - \mathbb{P}\big(R_N(i,t_i) > \varepsilon(i,t_i) \text{ for some } i \in \mathcal{I} \text{ and } t_i \in T_i\big) \\
&\qquad \geq 1 - \sum_{i \in \mathcal{I}} \sum_{t_i \in T_i} \mathbb{P}\big(R_N(i,t_i) > \varepsilon(i,t_i)\big)
\end{aligned} \tag{22}$$

$$\geq 1 - \sum_{i \in \mathcal{I}} \sum_{t_i \in T_i} \overline{\alpha}$$
$$= 1 - \overline{\alpha} \sum_{i \in \mathcal{I}} |T_i|,$$

where: the first equality follows from the probability of a complement; the first inequality follows from Boole's inequality; the second inequality follows from equivalence (21).

Therefore, for all $\alpha \in (0,1)$ and $N \in \mathbb{N}$, we have

$$\begin{aligned}\mathbb{P}\Big(\lambda \in \widehat{\Lambda}_N(\underline{\boldsymbol{\varepsilon}}_{\alpha,N})\Big) \geq 1-\alpha &\iff \mathbb{P}(R_N(i,t_i) \leq \varepsilon_{\alpha,N}(i,t_i;\lambda) \text{ for all } i \in \mathcal{I} \text{ and } t_i \in T_i) \geq 1-\alpha \\ &\impliedby \overline{\alpha} = \frac{\alpha}{\sum_{i\in\mathcal{I}} |T_i|} \\ &\iff \varepsilon_{\alpha,N}(i,t_i;\lambda) = \frac{\Big[K(i,t_i;\lambda)\sqrt{\ln|A_i|}\sqrt{\phi(t_i;\lambda)}\Big]\big[\sum_{i\in\mathcal{I}}|T_i|\big]}{\alpha\sqrt{N}},\end{aligned}$$

where: the first equivalence follows from the equivalences in (14); the implication follows from the chain of equalities and inequalities (22); the second equivalence follows from the right-hand side of equivalence (21). The desired result follows. ■

A.4 Proof of Theorem 4

Fix any $\varepsilon > 0$, and let $B^{|A|} := \{b \in \mathbb{R}^{|A|} : b^\intercal b \leq 1\}$ be the closed unit ball centered at $0_{|A|} \in \mathbb{R}^{|A|}$. First, note that

$$\lambda \in \widehat{\Lambda}_N(\varepsilon) \iff q_N \in Q(\varepsilon;\lambda).$$

Second, since $Q(\varepsilon;\lambda)$ is a non-empty, closed, and convex set, it has the following (support-function) characterization:

$$q_N \in Q(\varepsilon;\lambda) \iff \left\langle b^\intercal q_N - \sup_{q\in Q(\varepsilon;\lambda)} b^\intercal q \leq 0 \ \text{ for all } b \in B^{|A|}\right\rangle. \tag{23}$$

Third, since $Q(\varepsilon;\lambda)$ is also bounded and $b^\intercal q$ is continuous, the equivalence (23) reads as

$$q_N \in Q(\varepsilon;\lambda) \iff \left\langle b^\intercal q_N - \max_{q\in Q(\varepsilon;\lambda)} b^\intercal q \leq 0 \ \text{ for all } b \in B^{|A|}\right\rangle. \tag{24}$$

Fourth, since $b^\intercal q_N - \max_{q\in Q(\varepsilon;\lambda)} b^\intercal q = \min_{q\in Q(\varepsilon;\lambda)} b^\intercal(q_N - q)$ and $\min_{q\in Q(\varepsilon;\lambda)} b^\intercal(q_N-q)$ evaluated at $b = 0_{|A|} \in B^{|A|}$ is equal to 0, the equivalence (24) is equivalent to

$$q_N \in Q(\varepsilon;\lambda) \iff \left\langle \max_{b\in B^{|A|}} \min_{q\in Q(\varepsilon;\lambda)} b^\intercal(q_N - q) = 0\right\rangle. \tag{25}$$

Finally, note that: (i) the constraint that $b \in B^{|A|}$ can be written in terms of the first inequality constraints in program (P1); the constraint that $q \in Q(\varepsilon; \lambda)$ can be written in terms of the equality constraints and the last inequality constraint in program (P1); hence, the right-hand side of equivalence (25) is equivalent to $g(\lambda; q_N, \varepsilon) = 0$. The desired result follows. ■

B More General Environments

In this Appendix, we generalize the model in the main text by relaxing the assumption that payoff states are i.i.d. across periods. First, we present general dynamic environments. Next, we introduce ANR and study convergence to an appropriate BCCE notion. Finally, we discuss how our econometric approach extends to these more general environments.

B.1 General Dynamic Environments

When relaxing the assumption that payoff states θ_n are i.i.d. across periods, we must preserve meaningful notions of regret and convergence under ANR to some well-defined BCCE notion. Moreover, the limiting model must return helpful restrictions for estimating structural primitives. To achieve generality within these constraints, we split the vector of payoff states into two components. The first component, θ_{s}, relates to the "structural" features of the environment, which are invariant to counterfactuals. To fix ideas, these could be marginal costs in a pricing game or valuations in an auction. The other component, θ_{u}, relates to "non-structural" features that affect payoffs in the current environment but are not necessarily the same in a counterfactual. For instance, this component could include demand states that would change, e.g., if a digital platform would change format or market structure. We relax distributional assumptions as follows: we assume that "structural" payoff states have a limiting distribution without necessarily being i.i.d. across periods; we do not restrict the evolution of "non-structural" payoff states over time.

Formally, in a general dynamic environment, finitely many agents, $i \in \mathcal{I}$, interact over discrete-time periods, $n \in \mathbb{N}$.

- In each period n:
  - (i) Payoff states $\theta_{\mathrm{s},n}$ and $\theta_{\mathrm{u},n}$ are drawn from some state distributions $\psi_{\mathrm{s},n} \in \Delta\big(\overline{\Theta}_{\mathrm{s}}\big)$ and $\psi_{\mathrm{u},n} \in \Delta(\Theta_{\mathrm{u}})$, where $\overline{\Theta}_{\mathrm{s}}$ and Θ_{u} are the finite sets of payoff states.
  - (ii) Signals $t_n \coloneqq (t_{1,n}, \ldots, t_{I,n}) \in T \coloneqq T_1 \times \cdots \times T_I$ is drawn from some signal distribution $\overline{\pi}(\cdot \mid \theta_{\mathrm{s},n}) \in \Delta_{++}(T)$. Each agent i privately observes his signal $t_{i,n}$.
  - (iii) Each agent i selects an action $a_{i,n} \in A_i$.
  - (iv) Each agent i observes his realized payoff $\overline{u}_i(a_n, \theta_{\mathrm{s},n}, \theta_{\mathrm{u},n})$.

- Almost all paths of the process for payoff states $(\theta_{\mathrm{s},n})_{n=1}^{\infty}$ have a limiting empirical distribution $\psi_{\mathrm{s}} \in \Delta_{++}(\Theta_{\mathrm{s}})$ for some $\Theta_{\mathrm{s}} \subseteq \overline{\Theta}_{\mathrm{s}}$.

In this setting, Θ_{s} is the set of payoff states that occur infinitely often. Let π be the restriction of $\overline{\pi}$ to Θ_{s}. For all $i \in \mathcal{I}$, let u_i be the restriction of $\overline{u}_i$ to $A \times \Theta_{\mathrm{s}} \times \Theta_{\mathrm{u}}$. Under these assumptions, to the general dynamic environment $\mathcal{G}^{\infty}$, there corresponds a static game $\mathcal{G}^* \coloneqq (\mathcal{I}, G, S)$, where $G \coloneqq \big(\Theta_{\mathrm{s}}, \Theta_{\mathrm{u}}, (A_i, u_i)_{i=1}^{I}, \psi_{\mathrm{s}}\big)$ is the basic game and $S \coloneqq \big((T_i)_{i=1}^{I}, \pi_{\mathrm{s}}\big)$ is the information structure. In this section, we refer to $\mathcal{G}^*$ as the *limiting stage game*.

Discussion. To fix ideas, we note the following.

- General dynamic environments allow for many different stochastic processes governing the evolution of payoff states $(\theta_{\mathrm{s},n})_{n=1}^{\infty}$. Examples are i.i.d. payoff states, a perfectly persistent payoff state, and payoff states that follow various kinds of Markov chains.

- In the model we analyze throughout the main text (before our empirical application), we have $\theta_{\mathrm{s},n} = \theta_n$, and there is no $\theta_{\mathrm{u},n}$. Hence, $\Theta_{\mathrm{s}} = \overline{\Theta}_{\mathrm{s}} = \Theta$, and $\psi_{\mathrm{s}} = \psi$, where $\psi \in \Delta_{++}(\Theta)$ is the distribution from which payoff states are i.i.d. draws.

- In the empirical model in Section 5, we have $\theta_{\mathrm{s},n} = \zeta_{i,n}$ and $\theta_{\mathrm{u},n} = \overline{\rho}_{-i,n}$. Hence, $\Theta_{\mathrm{s}} = \overline{\Theta}_{\mathrm{s}} = T$, and $\Theta_{\mathrm{u}} = \Theta$. Moreover, $\psi_{\mathrm{s}} = \psi_i$, where $\psi_i \in \Delta_{++}(T)$ is the distribution from which seller i's marginal costs are i.i.d. draws, and $\pi\big(\zeta_i \,|\, \tilde{\zeta}_i\big) = 1$ if and only if $\zeta_i = \tilde{\zeta}_i$. Finally, payoff states $\theta_{\mathrm{u},n} = \overline{\rho}_{-i,n}$ are i.i.d. across periods, drawn from some distribution $\psi \in \Delta_{++}(\Theta)$.

As these observations show, general dynamic environments nest both our model in Sections 2–4 as well as our empirical model in Section 5.

Structural Parameters. The limiting stage game associated with the dynamic environment $\mathcal{G}^{\infty}$ belongs to a parametric class $\{\mathcal{G}^*(\lambda)\}_{\lambda \in \Lambda}$, indexed by the structural parameters λ.

B.2 Asymptotic No Regret

We extend the ANR notion to general dynamic environments as follows. For all $i \in \mathcal{I}$ and $t_i \in T_i$, the average factual payoff that agent i with signal t_i has obtained up to period N is

$$U_N(i, t_i) \coloneqq \frac{1}{N} \sum_{n=1}^{N} u_i\big((a_{i,n}, a_{-i,n}), \theta_{\mathrm{s},n}, \theta_{\mathrm{u},n}; \lambda\big) \mathbb{1}_{\{t_i\}}(t_{i,n}).$$

The average counterfactual payoff that agent i with signal t_i would have obtained had he played the best fixed action in hindsight up to period N is

$$V_N(i, t_i) \coloneqq \max_{a_i \in A_i} \left\{ \frac{1}{N} \sum_{n=1}^{N} u_i\big((a_i, a_{-i,n}), \theta_{\mathrm{s},n}, \theta_{\mathrm{u},n}; \lambda\big) \mathbb{1}_{\{t_i\}}(t_{i,n}) \right\}.$$

Regrets are defined as differences between these counterfactual and factual payoffs. That is, for all $i \in \mathcal{I}$ and $t_i \in T_i$, the regret of agent i with signal t_i before play in period $N+1$, denoted by $R_N(i, t_i)$, is defined as $R_N(i, t_i) \coloneqq \max\{V_N(i, t_i) - U_N(i, t_i), 0\}$.

Definition 12. *A sequence $(a_n, t_n, \theta_{\mathrm{s},n}, \theta_{\mathrm{u},n})_{n=1}^{\infty}$ from $\mathcal{G}^{\infty}$ has asymptotic no regret (ANR) if, for all $i \in \mathcal{I}$ and $t_i \in T_i$, we have $\limsup_{N\to\infty} R_N(i, t_i) \leq 0$ almost surely.*

B.3 Convergence under ANR

Bayes Coarse Correlated Equilibrium. A BCCE of the limiting stage game $\mathcal{G}^*(\lambda)$ is defined as follows.

Definition 13. *The probability distribution $\nu \in \Delta(A \times T \times \Theta_{\mathrm{s}} \times \Theta_{\mathrm{u}})$ is a Bayes coarse correlated equilibrium of $\mathcal{G}^*(\lambda)$ if:*

1. *ν is consistent for $\mathcal{G}^*(\lambda)$; that is, for all $t \in T$ and $\theta_{\mathrm{s}} \in \Theta_{\mathrm{s}}$, we have*

$$\sum_{a,\theta_{\mathrm{u}}} \nu(a, t, \theta_{\mathrm{s}}, \theta_{\mathrm{u}}) = \pi(t \mid \theta_{\mathrm{s}}; \lambda)\psi_{\mathrm{s}}(\theta_{\mathrm{s}}; \lambda).$$

2. *ν is coarsely obedient for $\mathcal{G}^*(\lambda)$; that is, for all $i \in \mathcal{I}$ and $t_i \in T_i$, we have*

$$\sum_{a,t_{-i},\theta_{\mathrm{s}},\theta_{\mathrm{u}}} \left[u_i\big((a'_i, a_{-i}), \theta_{\mathrm{s}}, \theta_{\mathrm{u}}; \lambda\big) - u_i(a, \theta_{\mathrm{s}}, \theta_{\mathrm{u}}; \lambda)\right]\nu(a, (t_i, t_{-i}), \theta_{\mathrm{s}}, \theta_{\mathrm{u}}) \leq 0 \quad \text{for all } a'_i \in A_i.$$

We denote by $E^*(\lambda)$ the set of BCCEs of $\mathcal{G}^*(\lambda)$.

ANR and Static Equilibria. The following proposition is the equivalent of (one of the two implications in) Theorem 1 in general dynamic environments.

Theorem 5 (Convergence under ANR)**.** *If the sequence $(a_n, t_n, \theta_{\mathrm{s},n}, \theta_{\mathrm{u},n})_{n=1}^{\infty}$ from $\mathcal{G}^{\infty}$ has ANR, then the sequence of empirical distributions $(Z_N)_{N=1}^{\infty}$ converges almost surely to $E^*(\lambda_0)$.*

Proof. Suppose the sequence $(a_n, t_n, \theta_{\mathrm{s},n}, \theta_{\mathrm{u},n})_{n=1}^{\infty}$ from $\mathcal{G}^{\infty}$ has ANR. Consider any subsequence $(Z_{N_l})_{l=1}^{\infty}$ of $(Z_N)_{N=1}^{\infty}$ that converges to some $\nu \in \Delta(A \times T \times \Theta_{\mathrm{s}} \times \Theta_{\mathrm{u}})$. We need to show that ν is almost surely consistent and coarsely obedient for $\mathcal{G}^*(\lambda_0)$.

Consistency. Pick any $(t, \theta_{\mathrm{s}}) \in T \times \Theta_{\mathrm{s}}$. Note that

$$\begin{aligned}\sum_{a,\theta_{\mathrm{u}}} \nu(a, t, \theta_{\mathrm{s}}, \theta_{\mathrm{u}}) &= \sum_{a,\theta_{\mathrm{u}}} \lim_{l\to\infty} Z_{N_l}(a, t, \theta_{\mathrm{s}}, \theta_{\mathrm{u}}) \\ &= \lim_{l\to\infty} \frac{\sum_{a,\theta_{\mathrm{u}}} Z_{N_l}(a, t, \theta_{\mathrm{s}}, \theta_{\mathrm{u}})}{\sum_{a,t,\theta_{\mathrm{u}}} Z_{N_l}(a, t, \theta_{\mathrm{s}}, \theta_{\mathrm{u}})} \lim_{l\to\infty} \sum_{a,t,\theta_{\mathrm{u}}} Z_{N_l}(a, t, \theta_{\mathrm{s}}, \theta_{\mathrm{u}}) \qquad (26)\end{aligned}$$

$$= \lim_{l\to\infty} \frac{\sum_{n=1}^{N_l} \mathbb{1}_{\{t\}}(t_n)\mathbb{1}_{\{\theta_\mathrm{s}\}}(\theta_{\mathrm{s},n})}{\sum_{n=1}^{N_l} \mathbb{1}_{\{\theta_\mathrm{s}\}}(\theta_{\mathrm{s},n})} \lim_{l\to\infty} \frac{\sum_{n=1}^{N_l} \mathbb{1}_{\{\theta_\mathrm{s}\}}(\theta_{\mathrm{s},n})}{N_l}.$$

The ratio

$$\frac{\sum_{n=1}^{N_l} \mathbb{1}_{\{t\}}(t_n)\mathbb{1}_{\{\theta_\mathrm{s}\}}(\theta_{\mathrm{s},n})}{\sum_{n=1}^{N_l} \mathbb{1}_{\{\theta_\mathrm{s}\}}(\theta_{\mathrm{s},n})} \tag{27}$$

is the empirical frequency of signal profile t at periods when the state is θ_s. Since signal profiles t_n are drawn from $\pi(\cdot \mid \theta_{\mathrm{s},n}; \lambda_0)$, the ratio in (27) is the empirical frequency of $\sum_{n=1}^{N_l} \mathbb{1}_{\{\theta_\mathrm{s}\}}(\theta_{\mathrm{s},n})$ conditionally independent draws from $\pi(\cdot \mid \theta_\mathrm{s}; \lambda_0)$. Moreover, since almost all paths of the process for payoff states $(\theta_{\mathrm{s},n})_{n=1}^\infty$ have a limiting empirical distribution ψ_s with full support on Θ_s, we have $\lim_{l\to\infty} \sum_{n=1}^{N_l} \mathbb{1}_{\{\theta_\mathrm{s}\}}(\theta_{\mathrm{s},n}) = \infty$ almost surely. Thus, by the strong law of large numbers,

$$\lim_{l\to\infty} \frac{\sum_{n=1}^{N_l} \mathbb{1}_{\{t\}}(t_n)\mathbb{1}_{\{\theta_\mathrm{s}\}}(\theta_{\mathrm{s},n})}{\sum_{n=1}^{N_l} \mathbb{1}_{\{\theta_\mathrm{s}\}}(\theta_{\mathrm{s},n})} = \pi(t \mid \theta_\mathrm{s}; \lambda_0) \quad \text{a.s.} \tag{28}$$

Moreover, again since almost all paths of the process for payoff states $(\theta_{\mathrm{s},n})_{n=1}^\infty$ have a limiting empirical distribution ψ_s with full support on Θ_s

$$\lim_{l\to\infty} \frac{\sum_{n=1}^{N_l} \mathbb{1}_{\{\theta_\mathrm{s}\}}(\theta_{\mathrm{s},n})}{N_l} = \psi_\mathrm{s}(\theta_\mathrm{s}; \lambda_0) \quad \text{a.s.} \tag{29}$$

From equations (26), (28), and (29), we have

$$\sum_{a,\theta_\mathrm{u}} \nu(a, t, \theta_\mathrm{s}, \theta_\mathrm{u}) = \pi(t \mid \theta_\mathrm{s}; \lambda_0)\psi_\mathrm{s}(\theta_\mathrm{s}; \lambda_0) \quad \text{a.s.} \tag{30}$$

We conclude from equation (30) that ν is almost surely consistent for $\mathcal{G}^*(\lambda_0)$.

Coarse obedience. Pick any $i \in \mathcal{I}$ and $t_i \in T_i$. For all $a_i' \in A_i$, let

$$V_N(i, t_i, a_i') \coloneqq \frac{1}{N} \sum_{n=1}^{N} u_i\big((a_i', a_{-i,n}), \theta_{\mathrm{s},n}, \theta_{\mathrm{u},n}; \lambda_0\big)\mathbb{1}_{\{t_i\}}(t_{i,n})$$

be the payoff of agent i with signal t_i for not having always played action a_i' up to period N. Note that

$$\begin{aligned} &V_N(i, t_i, a_i') - U_N(i, t_i) \\ &\quad = \frac{1}{N} \sum_{n=1}^{N} \big[u_i\big((a_i', a_{-i,n}), \theta_{\mathrm{s},n}, \theta_{\mathrm{u},n}; \lambda_0\big) - u_i\big((a_{i,n}, a_{-i,n}), \theta_{\mathrm{s},n}, \theta_{\mathrm{u},n}; \lambda_0\big)\big]\mathbb{1}_{\{t_i\}}(t_{i,n}) \\ &\quad = \sum_{a,t_{-i},\theta_\mathrm{s},\theta_\mathrm{u}} \big[u_i\big((a_i', a_{-i}), \theta_\mathrm{s}, \theta_\mathrm{u}; \lambda_0\big) - u_i(a, \theta_\mathrm{s}, \theta_\mathrm{u}; \lambda_0)\big] Z_N\big(a, (t_i, t_{-i}), \theta_\mathrm{s}, \theta_\mathrm{u}\big). \end{aligned} \tag{31}$$

Since $(a_n, t_n, \theta_{\mathrm{s},n}, \theta_{\mathrm{u},n})_{n=1}^{\infty}$ has ANR, for all $a_i' \in A_i$, we have

$$\limsup_{N\to\infty} \left[V_N(i, t_i, a_i') - U_N(i, t_i)\right] \leq 0 \quad \text{a.s.} \tag{32}$$

Then, by equation (31) and inequality (32), for all $a_i' \in A_i$, we have

$$\limsup_{N\to\infty} \sum_{a,t_{-i},\theta_{\mathrm{s}},\theta_{\mathrm{u}}} \left[u_i\big((a_i', a_{-i}), \theta_{\mathrm{s}}, \theta_{\mathrm{u}}; \lambda_0\big) - u_i(a, \theta_{\mathrm{s}}, \theta_{\mathrm{u}}; \lambda_0)\right] Z_N\big(a, (t_i, t_{-i}), \theta_{\mathrm{s}}, \theta_{\mathrm{u}}\big) \leq 0 \quad \text{a.s.} \tag{33}$$

Moreover, on the subsequence $(Z_{N_l})_{l=1}^{\infty}$, for all $a_i' \in A_i$, we have

$$\begin{aligned}
&\lim_{l\to\infty} \sum_{a,t_{-i},\theta_{\mathrm{s}},\theta_{\mathrm{u}}} \left[u_i\big((a_i', a_{-i}), \theta_{\mathrm{s}}, \theta_{\mathrm{u}}; \lambda_0\big) - u_i(a, \theta_{\mathrm{s}}, \theta_{\mathrm{u}}; \lambda_0)\right] Z_{N_l}\big(a, (t_i, t_{-i}), \theta_{\mathrm{s}}, \theta_{\mathrm{u}}\big) \\
&= \sum_{a,t_{-i},\theta_{\mathrm{s}},\theta_{\mathrm{u}}} \left[u_i\big((a_i', a_{-i}), \theta_{\mathrm{s}}, \theta_{\mathrm{u}}; \lambda_0\big) - u_i(a, \theta_{\mathrm{s}}, \theta_{\mathrm{u}}; \lambda_0)\right] \nu\big(a, (t_i, t_{-i}), \theta_{\mathrm{s}}, \theta_{\mathrm{u}}\big).
\end{aligned} \tag{34}$$

By inequality (33) and equation (34), for all $a_i' \in A_i$, we have

$$\sum_{a,t_{-i},\theta_{\mathrm{s}},\theta_{\mathrm{u}}} \left[u_i\big((a_i', a_{-i}), \theta_{\mathrm{s}}, \theta_{\mathrm{u}}; \lambda_0\big) - u_i(a, \theta_{\mathrm{s}}, \theta_{\mathrm{u}}; \lambda_0)\right] \nu\big(a, (t_i, t_{-i}), \theta_{\mathrm{s}}, \theta_{\mathrm{u}}\big) \leq 0 \quad \text{a.s.} \tag{35}$$

We conclude from equation (35) that ν is almost surely coarsely obedient for $\mathcal{G}^*(\lambda_0)$. ■

B.4 Econometrics

Given the convergence result in Theorem 5, we can develop an empirical strategy similar to that in Section 4 in the main text. In particular, using the BCCE restrictions in Definition 13, we can define analogous versions of the identified set and ε-BCCE estimator in Definitions 7 and 8 in the main text, and establish the analog to Theorem 2.

Developing an inferential procedure and establishing coverage properties, however, is not an immediate task. The confidence region we describe in Theorem 3 relies on the assumption that payoff states are i.i.d. across periods. Analogous results could be obtained under alternative assumptions on the law of the stochastic processes that may govern how payoff states evolve, as discussed in this appendix. Because making progress here relies on such additional assumptions, which are best motivated in the context of specific applications, we leave these extensions to future applications of our method.

C Internal Regrets and Bayes Correlated Equilibria

Another popular regret notion is that of internal regret (see, e.g., Cesa-Bianchi and Lugosi (2006); Nisan et al. (2007); Roughgarden (2016)), as opposed to external regret in Definition 1. In this appendix, we formalize the notions of asymptotic no internal regret and Bayes correlated equilibrium. Next, we study convergence to the set of Bayes correlated equilibria of the stage game under asymptotic no internal regret. Finally, we discuss the implications of asymptotic no internal regret on the econometric approach proposed by this paper.

C.1 Internal Regrets and Asymptotic No Internal Regret

Let a_i be the last action played by agent i with signal t_i before period N. For all pairs of actions $a_i, a_i' \in A_i$, let $V_i(a_i, a_i', t_i, N)$ be the average counterfactual payoff agent i with signal t_i would have obtained had he played a_i' every time up to period N he actually played a_i:

$$V_N(i, t_i, a_i, a_i') \coloneqq \frac{1}{N} \sum_{n=1}^{N} v_{i,n}(a_i, a_i', t_i),$$

where, for all $n \in \mathbb{N}$,

$$v_{i,n}(a_i, a_i', t_i) \coloneqq \begin{cases} u_i\big((a_i', a_{-i,n}), \theta_n; \lambda_0\big) \mathbb{1}_{\{t_i\}}(t_{i,n}) & \text{if } a_{i,n} = a_i \\ u_i\big((a_{i,n}, a_{-i,n}), \theta_n; \lambda_0\big) \mathbb{1}_{\{t_i\}}(t_{i,n}) & \text{if } a_{i,n} \neq a_i \end{cases}.$$

Definition 14 (Internal Regret). *For all $i \in \mathcal{I}$, $t_i \in T_i$, and $a_i, a_i' \in A_i$, the internal regret of agent i with signal t_i for action a_i' with respect to action a_i before play in period $N + 1$, denoted by $R_N(i, t_i, a_i, a_i')$, is defined as $R_N(i, t_i, a_i, a_i') \coloneqq \max\{V_N(i, t_i, a_i, a_i') - U_N(i, t_i), 0\}$.*

$R_N(i, t_i, a_i, a_i')$ is a measure of the time-average regret experienced by agent i with signal t_i at period N for not having played, every time that a_i was played in the past, the different action a_i'. When each agent has at most two actions, internal regrets coincide with regrets; otherwise, regrets are a coarser measure of regret than internal regrets.

Definition 15 (Asymptotic No Internal Regret). *A sequence $(a_n, t_n, \theta_n)_{n=1}^{\infty}$ from $\mathcal{G}^{\infty}$ has asymptotic no internal regret (ANIR) if, for all $i \in \mathcal{I}$, $t_i \in T_i$, and $a_i, a_i' \in A_i$, we have*

$$\limsup_{N \to \infty} R_i(a_i, a_i', t_i, N) \leq 0 \qquad \textit{almost surely.}$$

Let us refer to agent i with signal t_i as "agent (i, t_i)." ANIR requires the time average of the counterfactual increase in past payoffs, had each agent (i, t_i) changed each past play of a given action with its best replacement in hindsight, to vanish in the long run. In contrast, ANR requires the time average of the counterfactual increase in past payoffs, had each agent (i, t_i) played the best fixed action in hindsight, to vanish in the long run. Clearly, ANIR implies ANR; the converse is true only if each agent has at most two actions.

C.2 Bayes Correlated Equilibrium

Similarly to BCCE, a Bayes correlated equilibrium of the stage game $\mathcal{G}(\lambda)$ is a probability distribution over actions, signals, and states satisfying certain restrictions.

Definition 16 (Bayes Correlated Equilibrium)**.** *The probability distribution $\nu \in \Delta(A \times T \times \Theta)$ is a Bayes Correlated Equilibrium (BCE) of $\mathcal{G}(\lambda)$ if:*

1. *ν is consistent for $\mathcal{G}(\lambda)$ (see Definition 3).*
2. *ν is obedient for $\mathcal{G}(\lambda)$; that is, for all $i \in \mathcal{I}$, $t_i \in T_i$, and $a_i \in A_i$, we have*

$$\sum_{a_{-i}, t_{-i}, \theta} \left[u_i\big((a_i', a_{-i}), \theta; \lambda\big) - u_i\big((a_i, a_{-i}), \theta; \lambda\big)\right] \nu\big((a_i, a_{-i}), (t_i, t_{-i}), \theta\big) \leq 0 \quad \text{for all } a_i' \in A_i.$$

We denote by $\overline{E}(\lambda)$ the set of BCEs of $\mathcal{G}(\lambda)$ and by $\overline{Q}(\lambda)$ the set of BCE predictions of $\mathcal{G}(\lambda)$.

The BCE notion is due to Bergemann and Morris (2013, 2016) and differs from the BCCE notion in the form of incentive constraint—obedience *vs.* coarse obedience—we impose on ν. Suppose a mediator draws an action profile, a signal profile, and a state from distribution ν. Agents know ν. The mediator informs each agent i about his realized action-signal pair (a_i, t_i) from ν (and not only about his realized signal, as in a BCCE). A probability distribution ν is *obedient* if each agent i weakly prefers to play a_i, given that the other agents, who know their realized action-signal pair, play according to their part of the realized action profile.

The set $\overline{E}(\lambda)$ is convex. Since obedience is more demanding than coarse obedience, $\overline{E}(\lambda) \subseteq E(\lambda)$ and $\overline{Q}(\lambda) \subseteq Q(\lambda)$, i.e., the set of BCEs (resp., BCE predictions) is a subset of the set of BCCEs (resp., BCCE predictions). If each agent has at most two actions, coarse obedience and obedience coincide, and so $\overline{E}(\lambda) = E(\lambda)$ and $\overline{Q}(\lambda) = Q(\lambda)$. If $\mathcal{G}(\lambda)$ is a game with complete information, BCEs reduce to correlated equilibria (Aumann (1974, 1987)).

C.3 Convergence under Asymptotic No Internal Regret

The following theorem parallels Theorem 1 and shows that a sequence of actions, signals, and states from $\mathcal{G}^\infty$ has ANIR if and only if the sequence of empirical distributions of actions, signals, and states converges almost surely to the set of BCEs of $\mathcal{G}(\lambda_0)$.

Theorem 6 (Convergence under ANIR)**.** *The sequence $(a_n, t_n, \theta_n)_{n=1}^\infty$ from $\mathcal{G}^\infty$ has ANIR if and only if the sequence of empirical distributions $(Z_N)_{N=1}^\infty$ converges almost surely to $\overline{E}(\lambda_0)$.*

Proof. [$\Longrightarrow$] Suppose the sequence $(a_n, t_n, \theta_n)_{n=1}^\infty$ from $\mathcal{G}^\infty$ has ANIR. Consider any subsequence $(Z_{N_l})_{l=1}^\infty$ of $(Z_N)_{N=1}^\infty$ that converges to some $\nu \in \Delta(A \times T \times \Theta)$. We need to show that ν is almost surely consistent and obedient for $\mathcal{G}(\lambda_0)$.

Consistency. The proof of consistency is the same as that for Theorem 1.

Obedience. Pick any $i \in \mathcal{I}$ and $t_i \in T_i$. For all $a_i, a_i' \in A_i$, note that

$$V_N(i, t_i, a_i, a_i') - U_N(i, t_i)$$

$$= \frac{1}{N}\sum_{n=1}^{N}\left[u_i\big((a_i', a_{-i,n}), \theta_n; \lambda_0\big) - u_i\big((a_i^n, a_{-i,n}), \theta_n; \lambda_0\big)\right]\mathbb{1}_{\{a_i\}}(a_{i,n})\mathbb{1}_{\{t_i\}}(t_{i,n}) \tag{36}$$
$$= \sum_{a_{-i},t_{-i},\theta}\left[u_i\big((a_i', a_{-i}), \theta; \lambda_0\big) - u_i\big((a_i, a_{-i}), \theta; \lambda_0\big)\right]Z_N\big((a_i, a_{-i}), (t_i, t_{-i}), \theta\big).$$

Since $(a_n, t_n, \theta_n)_{n=1}^{\infty}$ has ANIR, for all $a_i, a_i' \in A_i$, we have

$$\limsup_{N\to\infty}\left[V_N(i, t_i, a_i, a_i') - U_N(i, t_i)\right] \leq 0 \quad \text{a.s.} \tag{37}$$

Then, by equation (36) and inequality (37), for all $a_i, a_i' \in A_i$, we have

$$\limsup_{N\to\infty}\sum_{a_{-i},t_{-i},\theta}\left[u_i\big((a_i', a_{-i}), \theta; \lambda_0\big) - u_i\big((a_i, a_{-i}), \theta; \lambda_0\big)\right]Z_N\big((a_i, a_{-i}), (t_i, t_{-i}), \theta\big) \leq 0 \quad \text{a.s.} \tag{38}$$

Moreover, on the subsequence $(Z_{N_l})_{l=1}^{\infty}$, for all $a_i, a_i' \in A_i$, we have

$$\lim_{l\to\infty}\sum_{a_{-i},t_{-i},\theta}\left[u_i\big((a_i', a_{-i}), \theta; \lambda_0\big) - u_i\big((a_i, a_{-i}), \theta; \lambda_0\big)\right]Z_{N_l}\big((a_i, a_{-i}), (t_i, t_{-i}), \theta\big)$$
$$= \sum_{a_{-i},t_{-i},\theta}\left[u_i\big((a_i', a_{-i}), \theta; \lambda_0\big) - u_i\big((a_i, a_{-i}), \theta\big); \lambda_0\right]\nu\big((a_i, a_{-i}), (t_i, t_{-i}), \theta\big). \tag{39}$$

By inequality (38) and equality (39), for all $a_i, a_i' \in A_i$, we have

$$\sum_{a_{-i},t_{-i},\theta}\left[u_i\big((a_i', a_{-i}), \theta; \lambda_0\big) - u_i\big((a_i, a_{-i}), \theta; \lambda_0\big)\right]\nu\big((a_i, a_{-i}), (t_i, t_{-i}), \theta\big) \leq 0 \quad \text{a.s.} \tag{40}$$

We conclude from equation (40) that ν is almost surely obedient for $\mathcal{G}(\lambda_0)$.

[$\Longleftarrow$] Now suppose $(Z_N)_{N=1}^{\infty}$ converges almost surely to $\overline{E}(\lambda_0)$. Pick any $i \in \mathcal{I}$ and $t_i \in T_i$. By obedience, for all $a_i, a_i' \in A_i$, we have

$$\limsup_{N\to\infty}\sum_{a_{-i},t_{-i},\theta}\left[u_i\big((a_i', a_{-i}), \theta; \lambda_0\big) - u_i\big((a_i, a_{-i}), \theta; \lambda_0\big)\right]Z_N\big((a_i, a_{-i}), (t_i, t_{-i}), \theta\big) \leq 0 \quad \text{a.s.} \tag{41}$$

By equation (36) and inequality (41), we have $\limsup_{N\to\infty}\left[V_N(i, t_i, a_i, a_i') - U_N(i, t_i)\right] \leq 0$ a.s. for all $a_i, a_i' \in A_i$, and so $\limsup_{N\to\infty} R_N(i, t_i, a_i, a_i') \leq 0$ a.s. The desired result follows.■

C.4 Estimation and Inference under ANIR

With the natural changes, one could implement the econometric method we propose by assuming that the sequence of actions, signals, and states from $\mathcal{G}^{\infty}$ satisfies ANIR (as opposed to ANR). In this case, the BCE notion would serve the role of BCCE in our main specification. Hence, BCEs would provide valid restrictions to estimate the structural parameters of interest under the ANIR assumption on behavior.

We developed our main results under ANR for two reasons. First, the two approaches are virtually equivalent from a conceptual viewpoint. Indeed, although ANIR is a more demanding benchmark metric to evaluate performance than ANR in online learning problems, Blum and Mansour (2007) show that there is always a "black box reduction" to convert any ANR algorithm into an ANIR algorithm. Thus, whenever agents can satisfy ANR, they can also satisfy ANIR (provided they perform the reduction of the algorithm). Second, assuming ANR is weaker (hence, more robust) than assuming ANIR. The reason is that ANR is a coarser objective than ANIR, and it does not require the additional assumption that agents can design an ANIR algorithm or are indeed converting an ANR algorithm into an ANIR algorithm.

Since $\overline{E}(\lambda) \subseteq E(\lambda)$ and $\overline{Q}(\lambda) \subseteq Q(\lambda)$, one may expect that implementing our econometric procedure under ANIR would yield more informative estimates. Indeed, defining an identified set using BCE restrictions under the ANIR assumptions would yield tighter restrictions. Similarly, for a fixed $\varepsilon > 0$, the ε-BCE estimator would result in a smaller estimated set of parameters than the ε-BCCE estimator.

In practice, however, the picture is different. Indeed, since ANIR is more demanding than ANR to satisfy, internal regrets converge to 0 more slowly than regrets. Hence, for given sample size, the $\underline{\boldsymbol{\varepsilon}}$ that would guarantee a desired coverage probability of the $\underline{\boldsymbol{\varepsilon}}$-BCE confidence region must be larger than the corresponding $\underline{\boldsymbol{\varepsilon}}$ that would guarantee the same coverage probability of the $\underline{\boldsymbol{\varepsilon}}$-BCCE confidence region. As a result, which approach would yield more informative estimates is unclear and may depend on the specific application. We leave exploring this comparison to future research.

D Additional Observables in the Data

D.1 Modeling Covariates

In this appendix, we describe how to incorporate exogenous observables into the method described in the main body of the paper. Suppose that also covariates $x \in X$ affect agents' payoffs. That is, each agent i's payoff in period n is $u_i(a_n, \theta_n, x)$—i.e., it is a function of the action profile $a_n \in A$ played in period n, the payoff state $\theta_n \in \Theta$ realized in period n, and the value of covariates $x \in X$ realized at the beginning of the game.

Covariates are realizations from the probability measure $\mathbb{P}_X$ with full support on some finite set X. We assume that X is finite to simplify the exposition; if X is not finite, one can similarly extend the model. Moreover, assume that θ_n is independent of x; this simplifying assumption is common in much of the literature on empirical games but can be relaxed.

Given a realization of the covariates x, the interaction unfolds as described in Section 2.1.

We denote the dynamic environment as $\mathcal{G}_x^\infty$ and the corresponding stage game as $\mathcal{G}_x(\lambda_0)$. We modify Assumption 1 as follows.

Assumption 2. *For all $x \in X$, consider a dynamic environment $\mathcal{G}_x^\infty$ with stage game $\mathcal{G}_x(\lambda_0)$:*

1. *For some positive integer N_x, the researcher observes an empirical distribution of actions $q_{N_x} \in \Delta(A)$ from $\mathcal{G}_x^\infty$.*
2. *The sequence of actions, signals, and states $(a_n, t_n, \theta_n)_{n=1}^\infty$ from $\mathcal{G}_x^\infty$ has ANR.*

We assume ANR for each realization of the covariates x, so the interaction is independent across different values of x. Intuitively, this corresponds to empirical environments where agents minimize regret across distinct markets and do not learn across them. However, some primitives are constant across markets—e.g., the distribution of marginal cost—while payoffs are shifted by covariates—e.g. because the demand system varies across markets.

We can now adapt the definitions of the identified set and confidence region. Let the probability measure $\mathbb{P}$ also capture the randomness in the covariates, and let $\mathbb{P}_x$ be the corresponding conditional probability measure given a realization of the covariates x. Since each stage game $\mathcal{G}_x(\lambda)$ has a corresponding set of BCCE predictions $Q_x(\lambda)$, we define the identified set as

$$\Lambda^* := \left\{ \lambda \in \Lambda : \mathbb{P}_x \left(\lim_{N_x \to \infty} d(q_{N_x}, Q_x(\lambda)) = 0 \right) = 1 \text{ for all } x \in X \right\}.$$

For all $x \in X$, let $\underline{\boldsymbol{\varepsilon}}_x := (\varepsilon_x(i, t_i))_{i \in \mathcal{I}, t_i \in T_i}$, where $\varepsilon_x(i, t_i) \geq 0$ for all $i \in \mathcal{I}$ and $t_i \in T_i$. Moreover, let $\underline{\boldsymbol{\varepsilon}} := (\underline{\boldsymbol{\varepsilon}}_x)_{x \in X}$ and $N := (N_x)_{x \in X}$. We define the $\underline{\boldsymbol{\varepsilon}}$-BCCE confidence region as

$$\widehat{\Lambda}_N(\underline{\boldsymbol{\varepsilon}}) := \{ \lambda \in \Lambda : q_{N_x} \in Q_x(\underline{\boldsymbol{\varepsilon}}; \lambda) \text{ for all } x \in X \}.$$

Clearly, the $\underline{\boldsymbol{\varepsilon}}$ required to establish coverage properties for the confidence region must now account for the different values of x.

We expect covariates to substantially increase the informativeness of estimates as they allow leveraging variation across different markets where agents interact.

D.2 Observing Actions Over Time

Suppose the researcher observes the empirical distribution of actions over time (or the action profile played in every period). In this case, Assumption 1 can be modified as follows.

Assumption 3. *Consider a dynamic environment $\mathcal{G}^\infty$ with stage game $\mathcal{G}(\lambda_0)$:*

1. *For some positive integer N, the researcher observes the empirical distribution of actions $q_n \in \Delta(A)$ for all $n = 1, \ldots, N$.*
2. *The sequence of actions, signals, and states $(a_n, t_n, \theta_n)_{n=1}^\infty$ from $\mathcal{G}^\infty$ has ANR.*

Under Assumption 3, the researcher can leverage the dynamic nature of the data-generating process for estimating the structural parameters λ_0. Since Assumption 3 differs from Assumption 1 about the observables (part 1) but not about agents' behavior (part 2), Theorem 1 continues to hold: almost surely, for all $\varepsilon > 0$, there exists $N_\varepsilon \in \mathbb{N}$ such that $q_N \in Q(\lambda_0; \varepsilon)$ for all $N \geq N_\varepsilon$. Therefore, for all $\varepsilon > 0$ and $\lambda \in \Lambda^*$, where Λ^* is the identified set in Definition 7, there exists N_ε such that $\lambda \in \widehat{\Lambda}_N(\varepsilon)$ for all $N > N_\varepsilon$. Hence, computing the ε-BCCE estimator $\widehat{\Lambda}_N(\varepsilon)$ for all $N > N_\varepsilon$ would provide helpful restrictions to learn about λ_0.

Based on these observations, if the researcher observes the empirical distribution of actions $q_n \in \Delta(A)$ for all $n = 1, \dots, N$, she can construct ε-BCCE estimators $\widehat{\Lambda}_n(\varepsilon)$ for successive values of n, and then consider their intersection. Formally, for some positive integer $K \leq N$, she can construct the *intersection of ε-BCCE estimators* of the identified set Λ^*:

$$\underline{\Lambda}_{K,N}(\varepsilon) \coloneqq \bigcap_{n=K}^{N} \widehat{\Lambda}_n(\varepsilon).$$

Clearly, for all $\varepsilon > 0$ and $K < N$, $\underline{\Lambda}_{K,N}(\varepsilon) \subseteq \widehat{\Lambda}_N(\varepsilon)$. Moreover, the theoretical convergence result in Theorem 1 implies that the following statement holds almost surely: for all $\varepsilon > 0$ and $\lambda \in \Lambda^*$, there exists N_ε such that

$$\lambda \in \bigcap_{n=N_\varepsilon}^{\infty} \widehat{\Lambda}_n(\varepsilon).$$

Thus, the intersection of ε-BCCE estimators, for any $\varepsilon > 0$, will recover in the limit valid bounds on λ_0. On the one hand, taking the intersection of ε-BCCE estimators can result in smaller bounds on structural parameters while retaining asymptotic almost-sure coverage properties similar to those of ε-BCCE estimators in Theorem 2. On the other hand, translating these ideas into a practical procedure raises several challenges. We mention here two.

First, given a sample size N: (i) how should we pick K? (ii) given a value of K, should we consider the intersection of all estimators for $n = K, \dots, N$, or "skip" some? Second, how should we set ε in the construction of $\underline{\Lambda}_{K,N}(\varepsilon)$? In the inferential procedure in the main text, we leverage theoretical convergence results for ANR algorithms to construct confidence regions with desired coverage properties. Extending this procedure to construct confidence regions based on the intersection of ε-BCCE estimators $\underline{\Lambda}_{K,N}(\varepsilon)$ requires much larger sample sizes (for reasonable coverage properties) and raises conceptual challenges.

Fully addressing these issues is outside the scope of this paper, and we leave further inquiry into a feasible inferential procedure based on $\underline{\Lambda}_{K,N}(\varepsilon)$ to future research.

E Computational Appendix

We discuss the computation of program (P1) in Section 4.4 in the context of our illustration, where the stage game has independent private values: $\Theta = T$, $u_i(a, \theta; \lambda) = u_i(a, t_i; \lambda)$, and π is such that $\pi(t \mid \theta; \lambda_0) = 1$ if and only if $t = \theta$ (hence, we can omit π from the program). However, the computational procedure also applies to the general setting of Sections 2–4. For our illustration, program (P1) becomes

$$\max_{b \in \mathbb{R}^{|A|}} \min_{\substack{q \in \mathbb{R}_+^{|A|} \\ \nu \in \mathbb{R}_+^{|A||T|}}} b^\intercal (q_N - q) \tag{P2}$$

subject to

$$\begin{aligned}
b^\intercal b - 1 &\leq 0, \\
q(a) - \sum_t \nu(a, t) &= 0 \quad \text{for all } a \in A, \\
\sum_a \nu(a, t) - \psi(t; \lambda) &= 0 \quad \text{for all } t \in T, \\
\sum_{a,t} \nu(a, t) - 1 &= 0, \\
\sum_{a, t_{-i}} \big[u_i(a_i', a_{-i}, t_i; \lambda) - u_i(a, t_i; \lambda)\big] \nu(a, t_i, t_{-i}) - \varepsilon &\leq 0 \quad \text{for all } i \in \mathcal{I}, a_i' \in A_i, t_i \in T_i.
\end{aligned}$$

The first, second, third, and fifth constraints correspond to the constraints in program (P1) in the context of our illustration. The fourth constraint is implied by the third constraint and the fact that $\psi(\cdot; \lambda)$ is a probability distribution; we add it explicitly to help computation.

Vectorization. Since the objective and all constraints in program (P2) are linear, we write them in matrix form. We represent the probability distribution$\psi(\cdot; \lambda)$ as a $|A| \times 1$ vector with elements $\psi(a_j; \lambda)$ for $1 \leq j \leq |A|$; analogously, we represent the probability distributions q and q_N as $|A| \times 1$ vectors. We represent a probability distribution $\nu \in \Delta(A \times T)$ as an $|A| \times |T|$ matrix with entries $\nu(a_j, t_k)$ for $1 \leq j \leq |A|$ and $1 \leq k \leq |T|$. Let $v \coloneqq \mathrm{vec}(\nu)$ be the vectorization of the matrix representation of ν; the linear transformation $\mathrm{vec}(\nu)$ stacks the columns of the matrix representation of ν on top on one other to obtain the $d_v \times 1$ vector v, where $d_v \coloneqq |A||T|$. Moreover, we define $z_1 \coloneqq q_N - q$, $z_2 \coloneqq v$, and the $d_z \times 1$ vector

$$z \coloneqq \begin{bmatrix} z_1 \\ z_2 \end{bmatrix},$$

where $d_z \coloneqq |A|(1 + |T|)$. We denote by 0_d the $d \times 1$ vector whose components are all zeros.

We write the equality constraints in program (P2) in matrix form as

$$M_{\text{eq}} z = y,$$

where: (i) M_{eq} is the appropriately defined $d_{\text{eq}} \times d_z$ matrix of coefficients, with $d_{\text{eq}} := |A| + |T| + 1$; (ii) y is the $d_{\text{eq}} \times 1$ vector defined as

$$y := \begin{bmatrix} q_N \\ \psi(\cdot;\lambda) \\ 1 \end{bmatrix}.$$

We write the inequality constraints in program (P2) in matrix form as

$$M_{\text{ineq}} z \leq \boldsymbol{\varepsilon},$$

where: (i) M_{ineq} is the appropriately defined $d_{\text{ineq}} \times d_z$ matrix of coefficients, with $d_{\text{ineq}} := \sum_i (|A_i||T_i|)$; (ii) $\boldsymbol{\varepsilon}$ is the $d_{\text{ineq}} \times 1$ vector whose components are all equal to ε. Finally, since $Q(\lambda;\varepsilon)$ is a subset of the $(|A|-1)$-dimensional simplex,

$$\max_{b \in B^{|A|}} \min_{q \in Q(\lambda;\varepsilon)} b^{\intercal}(q_N - q) = \max_{\tilde{b} \in B^{|A|-1}} \min_{q \in Q(\lambda;\varepsilon)} \begin{bmatrix} \tilde{b} \\ 0 \end{bmatrix}^{\intercal} (q_N - q),$$

where $B^{|A|-1}$ is the closed unit ball centered at $0_{|A|-1} \in \mathbb{R}^{|A|-1}$.

Therefore, we can now rewrite program (P2) in the following equivalent form:

$$\max_{\tilde{b} \in \mathbb{R}^{|A|-1}} \min_{\substack{z_1 \in \mathbb{R}^{|A|} \\ z_2 \in \mathbb{R}_+^{d_v}}} \begin{bmatrix} \tilde{b} \\ 0_{d_v+1} \end{bmatrix}^{\intercal} z \tag{P3}$$

subject to

$$\begin{aligned} \tilde{b}^{\intercal}\tilde{b} &\leq 1, \\ M_{\text{eq}} z &= y, \\ M_{\text{ineq}} z &\leq \boldsymbol{\varepsilon}. \end{aligned}$$

Duality. We replace the inner linear constrained minimization problem in program (P3) by its dual to obtain the following linear constrained maximization problem:

$$\max_{\substack{\tilde{b} \in \mathbb{R}^{|A|-1} \\ \ell_{\text{eq}} \in \mathbb{R}^{d_{\text{eq}}} \\ \ell_{\text{ineq}} \in \mathbb{R}_+^{d_{\text{ineq}}}}} - \begin{bmatrix} y \\ \boldsymbol{\varepsilon} \end{bmatrix}^{\intercal} \begin{bmatrix} \ell_{\text{eq}} \\ \ell_{\text{ineq}} \end{bmatrix} \tag{P4}$$

subject to

$$\tilde{b}^\intercal \tilde{b} \leq 1,$$

$$[M^\intercal]_{1:|A|} \begin{bmatrix} \ell_{\text{eq}} \\ \ell_{\text{ineq}} \end{bmatrix} = - \begin{bmatrix} \tilde{b} \\ 0 \end{bmatrix},$$

$$[M^\intercal]_{(|A|+1):d_z} \begin{bmatrix} \ell_{\text{eq}} \\ \ell_{\text{ineq}} \end{bmatrix} \geq 0_{d_v},$$

where

$$M \coloneqq \begin{bmatrix} M_{\text{eq}} \\ M_{\text{ineq}} \end{bmatrix}.$$

The dual variables associated to the constraints in program (P3) are the $d_{\text{eq}} \times 1$ vector ℓ_{eq} and the $d_{\text{ineq}} \times 1$ vector ℓ_{ineq}. $[M^\intercal]_{r:s}$ is the matrix consisting of rows $r, r+1, \ldots, s$ of matrix $M^\intercal$. Let $d_M \coloneqq d_{\text{eq}} + d_{\text{ineq}}$ be the number of rows of M. By strong duality, programs (P4) and (P3) have the same value. Program (P4) can be efficiently computed using standard solvers.

E.1 Computational Details for Section 3.3

The neighborhoods of the set of BCCE predictions $Q(\lambda_0)$ we represent in Figures 2 and 3 in Section 3.3 corresponds to the sets of $\underline{\boldsymbol{\varepsilon}}$-BCCE predictions $Q(\underline{\boldsymbol{\varepsilon}}; \lambda_0)$ in Definition 9 obtained by appropriately specifying the definition to capture the parametrization of our illustration. When setting $\underline{\boldsymbol{\varepsilon}} \coloneqq (\varepsilon(i, t_i; \lambda_0))_{i \in \mathcal{I}, t_i \in T_i}$, where $\varepsilon(i, t_i; \lambda_0) \geq 0$ for all $i \in \mathcal{I}$ and $t_i \in T_i$, we introduce dependence of $\varepsilon(i, t_i; \lambda_0)$ on the game's maximum deviation payoffs and set of actions, similar to what we do in Theorem 3. In particular, for all $i \in \mathcal{I}$ and $t_i \in T_i$, we set

$$\varepsilon(i, t_i; \lambda_0) \coloneqq \tilde{\varepsilon} \psi(t_i; \lambda_0) \Big[K(i, t_i; \lambda_0) \sqrt{\ln |A_i|} \sqrt{\psi(t_i; \lambda_0)} \Big],$$

where $K(i, t_i; \lambda_0)$ is defined as in Section 4.3, and $\tilde{\varepsilon} \in \{0.025, 0.05, 0.1\}$ depending on the figure or panel. The value of $\varepsilon(i, t_i; \lambda_0)$ above differs from the one in Theorem 3 only because of the presence of the scaling factor $\tilde{\varepsilon}$, and for the extra $\psi(t_i; \lambda_0)$ term that replaces the dependence from the sample size N, the confidence level α, and $\big[\sum_{i \in \mathcal{I}} |T_i|\big]$.

To compute the corresponding sets of $\underline{\boldsymbol{\varepsilon}}$-BCCE predictions, we discretize the unit ball $B^{|A|}$ in a grid of 100 values. For each b in the discretized ball, we compute $q(b)$ as the arg max of the following program:

$$\max_{\substack{q \in \mathbb{R}_+^{|A|} \\ \nu \in \mathbb{R}_+^{|A||T|}}} \; b^\intercal q \tag{P5}$$

subject to

$$
\begin{aligned}
q(a) - \sum_t \nu(a,t) = 0 \quad & \text{for all } a \in A, \\
\sum_a \nu(a,t) - \psi(t;\lambda_0) = 0 \quad & \text{for all } t \in T, \\
\sum_{a,t} \nu(a,t) - 1 = 0, \quad & \\
\sum_{a,t_{-i}} \big[u_i(a_i', a_{-i}, t_i; \lambda_0) - u_i(a, t_i; \lambda_0)\big]\nu(a, t_i, t_{-i}) - \varepsilon(i, t_i; \lambda_0) \leq 0 \quad & \text{for all } i \in \mathcal{I}, a_i' \in A_i, t_i \in T_i.
\end{aligned}
$$

Program (P5) is linear and shares the BCCE constraints with program (P2). We solve it using KNITRO in AMPL. To draw Figures 2 and 3, we plot the convex hull of all $q(b)$ using Matlab's MPT Toolbox (Kvasnica, Grieder, and Baotić (2004)).

E.2 Computational Details for Simulations

For our simulations in Section 4.5, we adapt program (P4) by specifying the matrices M_{eq} an M_{ineq} to reflect the parametrization adopted for the set of types, utility function, and ψ. This results in a program with 901 variables and 101 equality constraints. We solve the program in the Matlab-based modeling system CVX (see Grant and Boyd (2008, 2014)) using the solver SDPT3. The evaluation of program (P4) for any candidate parameter value λ takes about 0.4 seconds with a 3.2GHz processor.

We construct the sets $\widehat{\Lambda}_N(\underline{\boldsymbol{\varepsilon}}_{\alpha,N})$ for our simulations by evaluating program (P4) on a grid of 10,000 parameter values. We report the projections of these sets for μ and σ in Figures 4 and 5, panel (b). To enhance the graphical representation of $\widehat{\Lambda}_N(\underline{\boldsymbol{\varepsilon}}_{\alpha,N})$ in the figures, we construct a larger grid of 500,000 parameters and classify them as in or out of $\widehat{\Lambda}_N(\underline{\boldsymbol{\varepsilon}}_{\alpha,N})$ by training a Support Vector Machine (SVM) on the 10,000 grid points where we evaluated program (P4). While this procedure produces figures with smoother boundaries, it does not alter the overall shape of the identified set in any way.

E.3 Computational Details for the Application

To compute the estimates in Section 5.4, we follow similar steps as those outlined above for our simulations. Specializing matrices M_{eq} and M_{ineq} yields a program with 111 variables and 36 equality constraints. Solving this version of program (P4) for any λ_i takes about 0.3 seconds with a 3.2GHz processor. To prevent numerical problems due to dividing by near-zero numbers, we implement $\Phi(i,\zeta_i) = \sum_{\rho_i \in A: \Phi(i,\rho_i,\zeta_i) > 0.1K(i,\zeta_i)} \frac{K(i,\zeta_i)}{\Phi(i,\rho_i,\zeta_i)}$.

As for our simulations, we construct the sets $\widehat{\Lambda}_{i,N}(\underline{\boldsymbol{\varepsilon}}_{\alpha,N})$ by evaluating the application-specific formulation of program (P4) on a grid of 10,000 parameter values. The projections of these sets for μ_i and σ_i are reported in Figure 7. We then construct a larger grid of 500,000 parameters and classify them by training SVM on the original 10,000 grid points.

F Empirical Application: Further Details

We provide supporting information and details for the empirical application in Section 5.

F.1 Dateset Construction and Description

We construct our Swappa data by focusing on iPhone models only and dropping all iPhone generations older than iPhone 8: these correspond to a small part of the data, since iPhone 8 devices are already outdated at the beginning of our sample period. We then drop some observations that correspond to price mistakes or are otherwise unrealistic. We drop all listings with a price greater than 120% of the new price for the corresponding device model on the Apple Store, prices greater than 200% of the Decluttr reference price, or prices below 50% of the Decluttr reference price for the device model. We also drop all prices that differ by more than $250 than the Decluttr reference price. Finally, we dropped all devices that were on sale for more than 100 days.

Table 3: Descriptive Statistics at the Device-Day Level.

	Mean	SD	P25	P75	Min	Max
Full sample (4,980 sellers, 259,102 obs.)						
Listing prices ($)	495.87	246.95	289.00	685.00	75.00	1,499.00
Reference prices ($)	523.37	250.90	319.99	709.99	109.99	1,329.99
Days before sold	9.61	13.37	2.00	12.00	0.00	99.00
Top 15 sellers (83,801 obs.)						
Listing prices ($)	471.85	234.27	276.00	650.00	80.00	1,333.00
Reference prices ($)	510.48	242.02	309.99	709.99	119.99	1,329.99
Days before sold	9.90	12.98	2.00	12.00	0.00	94.00
Most popular device gens. (56,965 obs.)						
Listing prices ($)	390.04	110.12	299.00	461.00	176.00	772.00
Reference prices ($)	408.93	101.29	339.99	459.99	259.99	729.99
Days before sold	7.41	11.19	1.00	9.00	0.00	99.00

We report descriptive statistics for our main variables at the device-day level. "Days before sold" is the number of days between an item's creation date and its sold date. Top sellers are those who sold the most devices during the sample period and actively listed devices every day. The three most popular device generations are iPhone 13, iPhone 12, and iPhone 11. *Source: Swappa data collected by the authors.*

These steps yield the device-day level database described in Table 3. The full dataset includes 259,102 observations we scraped from Swappa (listing prices and sales indicators) and matched with reference price data scraped from Decluttr. Across our sample, we observe substantial variation in both listing and reference prices; mean listing and reference prices are both around $500, with a standard deviation of about $250. The variation in listing prices stems both from the fact that we consider a range of different device models on sale and from heterogeneity in sellers' pricing behavior. On average, a device gets sold on Swappa about 10 days after being listed. We further explore both of these channels below. When restricting our data to the top-15 largest sellers (by sales), the statistics for listing prices and time to sale are broadly similar. When we look at the most popular device models on the platform, which are not the latest iPhone generations available during our sample period, these tend to be cheaper and sell faster.

In Table 4, we report descriptive statistics at the seller level. We observe 4,980 sellers in the data, although only about 417 (on average) have an active listing each day on the platform. Active sellers have (on average) 42 listings per day, make 4,571 daily pricing decisions over the sample period, and have sales for around $2,000. These averages, however, hide substantial asymmetry: the largest sellers are responsible for a large share of listings and sales. On average, the top-15 sellers have 84 devices on sale per day, make more than 10,000 daily pricing decisions over the sample periods, and have average revenues that exceed $100,000. Average revenues are more than three times as high if we look at the top-2 sellers.

Table 4: Descriptive Statistics at the Seller Level.

	Mean	SD	P25	P75	Min	Max
Full sample (4,980 sellers, 259,102 obs.)						
# of sellers (per day)	417	194	230	605	182	813
# of devices (per seller-day)	42	58	3	55	1	395
# of device-days (per seller)	4,571	5,658	181	7,187	1	18,750
Revenues from devices sold ($)	1,931	12,080	412	902	78	326,677
Top-15 sellers (83,801 obs.)						
# of devices (per seller-day)	82	66	33	114	1	331
# of device-days (per seller)	9,896	5,896	4,833	14,725	1,286	18,294
Revenues from devices sold ($)	95,770	88,684	44,170	122,090	23,171	306,677
Top-2 sellers (33,019 obs.)						
# of devices (per seller-day)	126	55	80	162	17	265
# of device-days (per seller)	16,702	1,774	14,725	18,294	14,725	18,294
Revenues from devices sold ($)	297,479	13,008	288,281	306,677	288,281	306,677

We report descriptive statistics at the seller level *Source: Swappa data collected by the authors.*

F.2 Large Sellers on Swappa

In Table 5, we report information on the largest sellers on Swappa. All of the top-15 sellers on the platform sell a considerable number of devices over our sample period (with either more than around 250 listings or more than around 100 devices sold). However, the top-2 sellers on the platform stand out in terms of the range of models and volume of devices they offer.

Table 5: Large Sellers on Swappa.

	Seller Company Name	Device-Day Obs.	# Listings	# Sold	# Models
Top-2 sellers					
	AMS Traders Inc.	18,294	1,512	620	23
	GadgetPickup	14,725	1,695	683	23
Other top-15 sellers					
	SaveGadget	8,025	658	243	20
	Certified Cell, Inc	7,697	904	237	18
	UPGRADE SOLUTION INC	7,173	903	427	3
	NYCPhonebuyer LLC	6,037	743	336	23
	Buyback Surgeon	4,833	286	113	21
	Brandon K.	4,381	323	133	18
	TD'S Cell Shop	2,301	316	164	15
	Cybertron Electronics LLC	2,198	234	108	7
	GreenStream	1,981	202	92	17
	Gilly's Smart Phones	1,875	388	190	12
	cell4pets	1,597	228	124	19
	Tech Plug Electronics	1,398	248	89	22
	Sellworld	1,286	255	100	16

We report top-15 seller names and statistics. *Source: Swappa data scraped by the authors.*

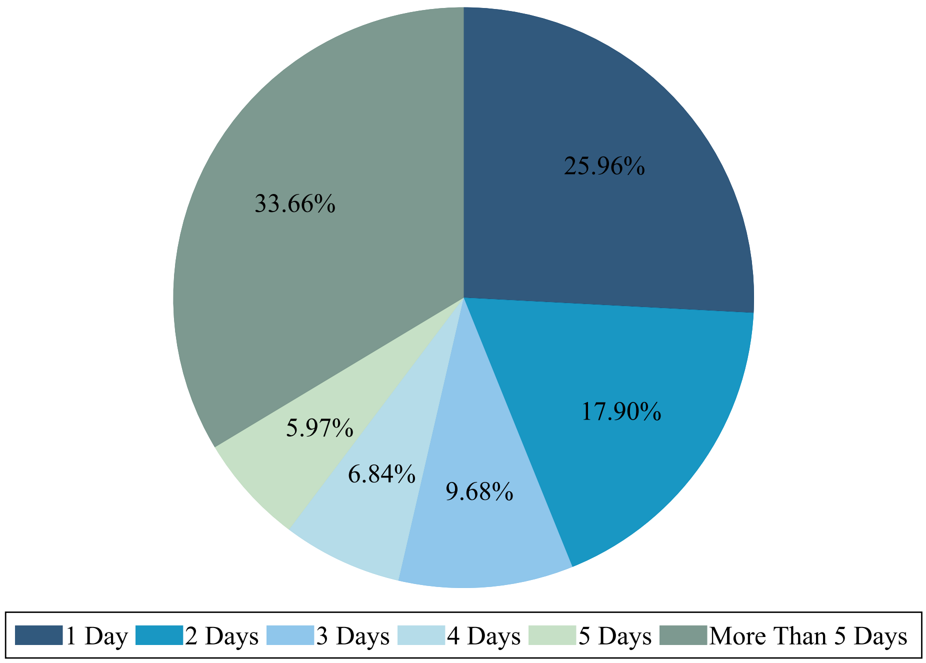
F.3 Daily Price Decision

We model sellers on Swappa as making daily pricing decisions; thus, we collect data with a daily frequency. This modeling decision reflects our understanding of the institutional context. To assess this assumption, we look at the frequency of price changes in the data, with an important caveat: our ANR assumption does not imply that sellers must change the price they charge *every period*, and thus allows for substantial inertia in pricing.

We considered a range of possible assumptions on the frequency of decision-making, ranging from intra-daily (e.g., every six or twelve hours) to daily, bi-weekly, or weekly. Although we experimented with higher-frequency data collection, we found few intra-day price changes, thus suggesting that sub-daily decision-making is not plausible. Analysis of daily data shows that 26% of price changes modify a price set the previous day, while 54% alter a price set within the last three days.[18] In Figure 8, we illustrate the complete distribution of price-changing frequencies. While not all prices change daily on Swappa, the daily price-setting model captures market dynamics more accurately than bi-weekly or weekly alternatives, which would obscure significant price variation in the data.

[18]These figures refer to the top-15 sellers on the platform; the values for the top-2 sellers are similar.

Figure 8: Price Changing Frequency (Top-15 Sellers).



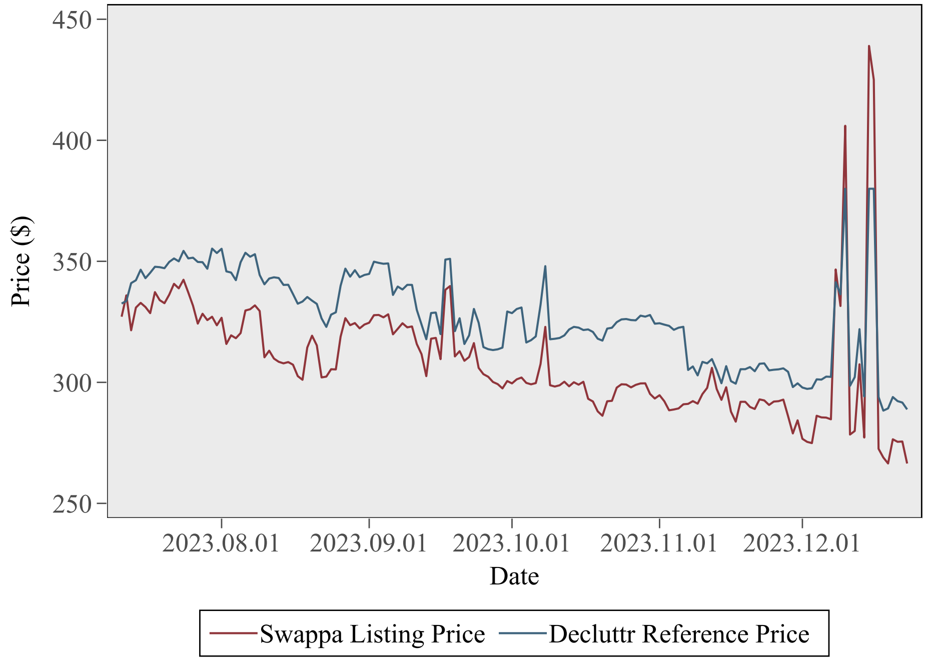
We represent the distribution of the frequency of price changes for the top-15 sellers on Swappa. For each price change $p_{i,n} \neq p_{i,n-1}$ (not considering the first pricing decision when a new device appears on the platform), we keep track of how long $p_{i,n-1}$ was unchanged. We then count the frequency of each time interval and plot it in the pie chart. *Source: Swappa data collected by the authors.*

F.4 Reference Pricing, Discretization, and Sale Probabilities

To ensure tractability of the model and leverage our data, we subtract for each device-day pair $n = (j, d)$ reference prices $\chi_{m(j),d}$ from Swappa prices $p_{i,n}$ to obtain residuals $\rho_{i,n}$. These residuals capture sellers' decisions on how to price their devices relative to the reference price in the broader market for used iPhone devices. Corresponding to this intuition, we show below that the data on residual prices contain useful variation to understand sellers' incentives.

First, to give insights on how Swappa listing prices and Decluttr reference prices co-vary over time, we plot in Figure 9 the paths of average prices for the most popular device generation in our data, iPhone 11. In line with the descriptive statistics in Appendix F.1, Swappa average listing prices are below reference prices, and track them closely. Reference prices capture seasonal trends (e.g., spikes around the holidays) that are also reflected in the Swappa prices. However, there is variation in average residual prices for this device generation, reflecting Swappa-specific supply and competitive conditions, e.g., sellers' strategies and costs. While we focus on one device in the plot, paths for other devices are qualitatively similar.

Figure 9: Swappa Listing Prices and Decluttr Reference Prices – iPhone 11.



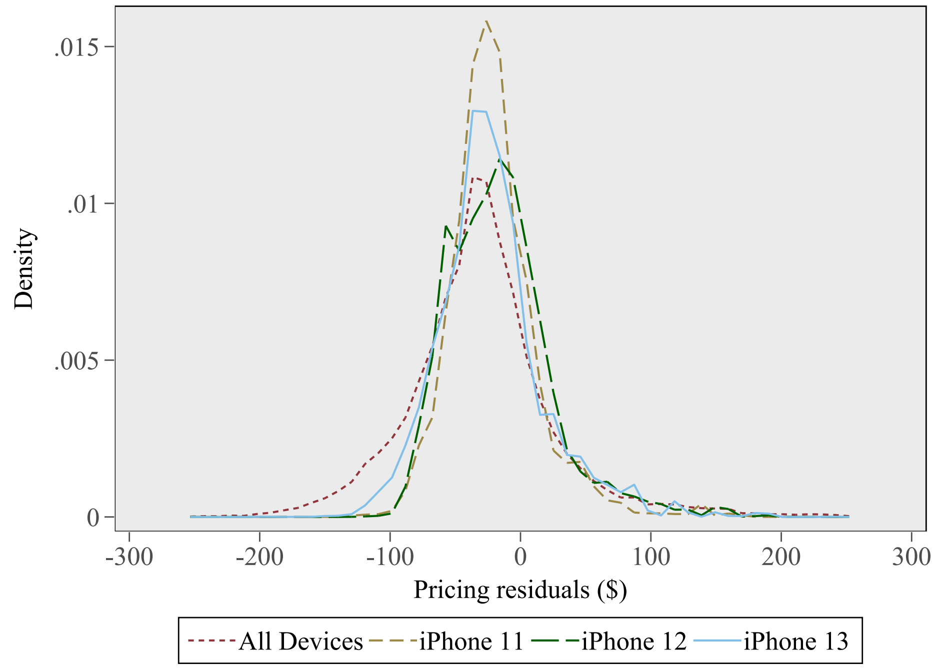
The figure represents average Swappa listing prices (in red) and Decluttr reference prices (in blue) for iPhone 11 devices over our sample period. *Source: Swappa and Decluttr data collected by the authors.*

Pricing residuals $\rho_{i,n}$ thus exhibit significant variation across time, devices, and sellers. We plot in Figure 10 the overall distribution of $\rho_{i,n}$ and the distributions for the three device generations that are most common in our data. All distributions are centered below 0, reflecting lower average prices on Swappa than on Decluttr. Despite some differences in shape, all distributions seem to follow a Normal pattern, with large deviations in either direction less common than small deviations. We take this as further evidence that reference prices, which reflect broader market conditions, anchor sellers' pricing decisions on Swappa meaningfully.

Because we capture discrete pricing decisions in our model, we discretize $\rho_{i,n}$ into five bins. To construct the bins, we first compute pricing residuals for *all* sellers (including non top-15 sellers) on the platform across the whole sample. From their distribution, we construct five quantiles and, therefore, five discrete "bins" corresponding to discrete levels of $\rho_{i,n}$. Although we use the data to construct the discretized set of actions, at this stage, we use observations from the entire set of sellers, while the estimation in Section 5.4 only focuses on the top-2 sellers. Figure 11 shows the variation in discretized $\rho_{i,n}$ for the top-2 sellers in the data. Even after discretization, the variation in pricing residuals across time and sellers is preserved.

A relevant primitive for our empirical model is a device's sale probability, $s(\rho_{i,n}, \bar{\rho}_{-i,n})$. Such a probability is a function of the seller's price residual $\rho_{i,n}$ and of the competitors'

Figure 10: Distribution of Pricing Residuals.



We represent the distribution of $\rho_{i,n}$ in the data. The red dashed line represents the overall distribution across all devices; the yellow, green, and blue lines represent the distributions for iPhones 11, 12, and 13.

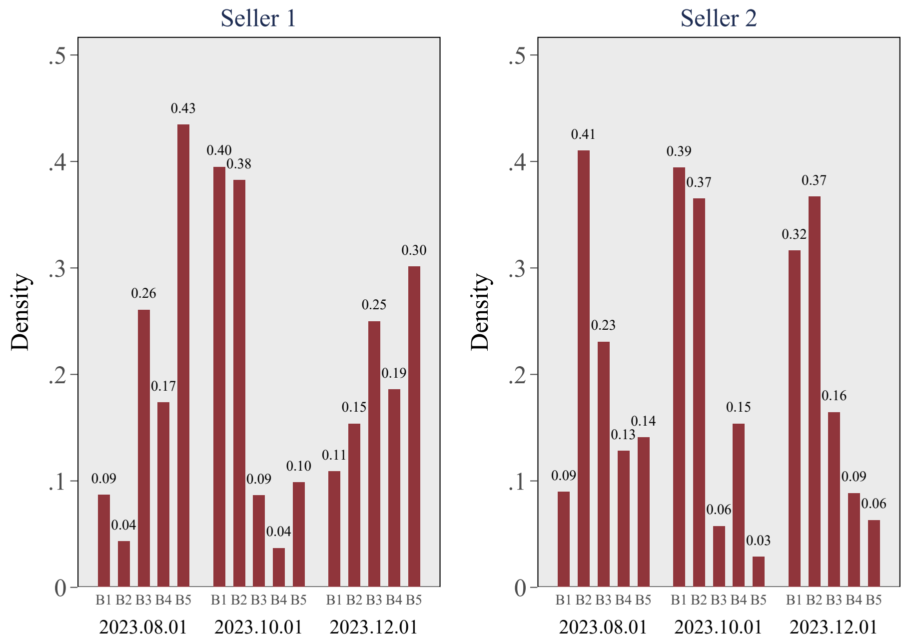
average price residual $\bar{\rho}_{-i,n}$. Because both $\rho_{i,n}$ and $\bar{\rho}_{-i,n}$ are discretized, this function takes on $|A|^2 = 25$ values; we estimate it from the data as the empirical sale probability for a device-day. We compute these probabilities using data from *all* sellers in our data. Underlying this approach there is an assumption that, by differencing out reference prices, we have appropriately controlled for all relevant device characteristics. Because of the fine level at which we match reference prices to devices, keeping track of device model, memory, and condition, we believe this assumption is appropriate in our context.

Figure 12 represents the estimated sale probabilities $s(\rho_{i,n}, \bar{\rho}_{-i,n})$ for each combination of $\rho_{i,n}$ and $\bar{\rho}_{-i,n}$. The sale probabilities range between 0.5 percent and 10 percent. In line with economic intuition, the probability of selling a device decreases with the own (discretized) price residual $\rho_{i,n}$ and increases with competitors' average price residual $\bar{\rho}_{-i,n}$. This supports the view that we are appropriately controlling for unobserved heterogeneity.

F.5 Gazelle Data

In our markup quantification exercise in Section 5.5, we use data from Gazelle, a centralized buyback and resale platform that directly purchases devices from consumers, refurbishes

Figure 11: Distribution of 5 Bin over Time (Top-15 Sellers).



We represent the distribution of the time-average of $\rho_{i,n}$, discretized into five bins, at three points in time in our sample. The two panels represent sellers 1 and 2, the largest in our data.

them, and resells them.[19] Gazelle's website is similar to Decluttr, where we collect our reference price data. However, Decluttr stopped buying used devices during our sample period, thus prompting us to collect data from Gazelle.[20]

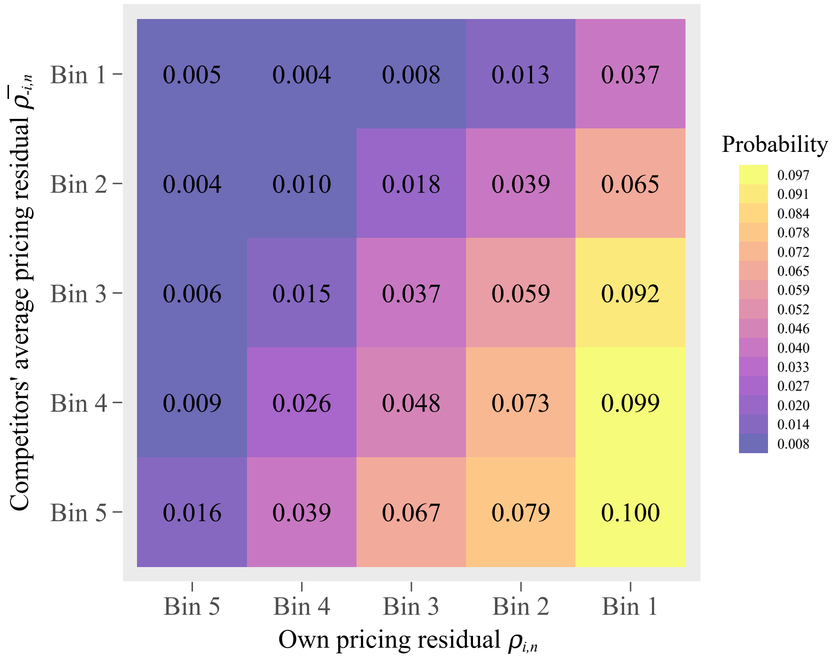
We scrape the website for 39 days, collecting prices at which the platform sells ("ask") and buys ("bid") devices of a given model (including storage capacity and condition). We average these prices across all daily observations in our sample and report descriptive statistics for the Gazelle data in Table 7. Across the entire dataset, we see average ask prices of about \$500, with significant variation across models and conditions. These values are close to what we observe for our Swappa and Decluttr databases in Appendix F.1. The average bid-ask spread on Gazelle is \$197, and markups are between \$108 and \$167. While variation across iPhone generations and storage is the most significant driver of heterogeneity, condition plays an important part. Devices in excellent condition command a premium and have higher average bid-ask spreads than devices in fair condition (around \$217 versus \$189).

Finally, we discuss our choice of estimating the costs that Gazelle incurs to market the devices it purchases at 10-30% of the bid price. Estimating the entity of these costs is chal-

[19]See `https://buy.gazelle.com/`.

[20]Since then, Decluttr has changed its strategy again to selling devices on eBay and Amazon but restarted accepting trade-ins. Gazelle continues to operate through its website.

Figure 12: Sale Probability.



We represent empirical sales probabilities $s(\rho_{i,n}, \bar{\rho}_{-i,n})$ for each discretized value of $\rho_{i,n}$ (on the horizontal axis, in bins 1–5) and of $\bar{\rho}_{-i,n}$ (on the vertical axis, in bins 1–5). Each cell reports the value of $s(\cdot,\cdot)$; darker colors correspond to lower sale probabilities.

lenging. Because Gazelle is not a public company, financial data are not readily available. We use a source[21] that quantifies industry estimates for different components of marginal cost. For an average device (e.g., for the average bid price of about \$300), a seller incurs costs for processing, testing, and quality assurance of \$15-65; warranty and customer support adds \$5-20 to the cost. Shipping (including the possibility of returns) adds about \$5. We do not add here refurbishment and repair costs (labor and parts), which are more substantial (\$25-100 per device), as Gazelle grades the device it receives according to the same scale of the device it sells, and so does not meaningfully refurbish the devices it resells. Overall, this exercise measures additional costs at around 10-30% of the bid price.

F.6 Empirical Model: Proofs and Technical Details

F.6.1 Proof of Proposition 1

Preliminaries: Regret and Asymptotic No Regret. We establish the formal notation by providing definitions of factual payoffs, counterfactual payoffs, regret, and asymptotic no

[21] See this link; accessed on September 2024.

Table 7: Bid and Ask Prices, Markups from Gazelle.

	Mean	SD	P25	P75	Min	Max
Full Dataset (237 observations)						
Sell/“Ask” price ($)	491.54	268.71	254.77	694.33	109.39	1,316.76
Buy/“Bid” price ($)	294.71	170.17	143.56	428.26	54.05	779.46
Bid-ask spread ($)	196.83	100.09	113.31	269.87	55.34	537.29
Markups ($)	[108.41, 167.35]					
Excellent (79 observations)						
Sell/“Ask” price ($)	517.77	281.50	270.63	747.97	135.28	1,316.76
Buy/“Bid” price ($)	309.15	176.23	151.62	457.66	59.41	779.46
Bid-ask spread ($)	208.61	107.07	118.87	291.96	72.23	537.29
Markups ($)	[115.87, 177.70]					
Good (79 observations)						
Sell/“Ask” Price ($)	491.76	267.05	257.66	695.31	119.36	1,260.28
Buy/“Bid” price ($)	294.17	169.93	143.24	428.26	57.77	758.29
Bid-ask spread ($)	197.59	98.55	115.67	277.05	61.41	501.99
Markups ($)	[109.34, 168.17]					
Fair (79 observations)						
Sell/“Ask” price ($)	465.09	257.87	237.11	663.85	109.39	1,212.36
Buy/“Bid” Price ($)	280.81	165.16	132.29	410.62	54.05	741.67
Bid-ask spread ($)	184.27	93.98	104.20	254.97	55.34	470.69
Markups ($)	[100.03, 156.19]					

We report prices and markups on Gazelle. Each observation (i.e., device) is defined by a combination of model (e.g., iPhone 15 Pro), storage capacity (e.g., 128GB), and condition (i.e., “Fair,” “Good,” or “Excellent”). Sell/“ask” price refers to the prices at which the platform sells to customers, while buy/“bid” price refers to the prices at which the platform buys from individuals. The bid-ask spread is the amount by which the ask exceeds the bid price. Markups are obtained by subtracting from Gazelle sell ask prices the bid prices times a cost multiplier. We consider multipliers of 1.1 and 1.3 to give a conservative interval. *Source: Gazelle data collected by the authors.*

regret in the context of our empirical model. Consider seller i. For all $\zeta_i \in T$, the average factual payoff that seller i with marginal cost ζ_i has obtained up to period N is

$$U_N(i,\zeta_i) \coloneqq \frac{1}{N}\sum_{n=1}^{N} s(\rho_{i,n},\overline{\rho}_{-i,n})(\rho_{i,n}-\zeta_{i,n})\mathbb{1}_{\{\zeta_i\}}(\zeta_{i,n}).$$

Moreover, the average counterfactual payoff that seller i with marginal cost ζ_i would have obtained had he always set price ρ'_i up to period N is

$$V_N(i,\zeta_i,\rho'_i) \coloneqq \frac{1}{N}\sum_{n=1}^{N} \left[s(\rho'_i,\overline{\rho}_{-i,n})(\rho'_i-\zeta_{i,n})\right]\mathbb{1}_{\{\zeta_i\}}(\zeta_{i,n}).$$

Let $V_N(i,\zeta_i) \coloneqq \max_{\rho'_i\in A} V_N(i,\zeta_i,\rho'_i)$. The regret of seller i with marginal cost ζ_i for not having always set the best fixed price in hindsight up to period N is $R_N(i,\zeta_i) \coloneqq \max\{V_N(i,\zeta_i) - U_N(i,\zeta_i),0\}$. The sequence of prices, marginal costs, and states $(\rho_{i,n},\zeta_{i,n},\overline{\rho}_{-i,n})_{n=1}^{\infty}$ from $\mathcal{G}_i^{\infty}$

has ANR if, for all $\zeta_i \in T$, we have $\limsup_{N\to\infty} R_N(i,\zeta_i) \leq 0$ almost surely.

Proof of Proposition 1. Consider any subsequence $(Z_{i,N_l})_{l=1}^{\infty}$ of $(Z_{i,N})_{N=1}^{\infty}$ that converges to some $\nu_i \in \Delta(A\times T\times\Theta)$, where $\nu_i = \tilde{\nu}_i \times \psi$ for some $\tilde{\nu}_i \in \Delta(A\times T)$. We need to show that $\tilde{\nu}_i$ is almost surely consistent and coarsely obedient for $\mathcal{G}_i(\lambda_{i,0})$ according to Definition 11.

Consistency. Pick any $\zeta_i \in T$. Note that

$$\begin{aligned}\sum_{\rho_i} \tilde{\nu}_i(\rho_i,\zeta_i) &= \sum_{\rho_i,\overline{\rho}_{-i}} \tilde{\nu}_i(\rho_i,\zeta_i)\psi(\overline{\rho}_{-i}) \\ &= \sum_{\rho_i,\overline{\rho}_{-i}} \nu_i(\rho_i,\zeta_i,\overline{\rho}_{-i}) \\ &= \sum_{\rho_i,\overline{\rho}_{-i}} \lim_{l\to\infty} Z_{i,N_l}(\rho_i,\zeta_i,\overline{\rho}_{-i}) \\ &= \lim_{l\to\infty} \frac{\sum_{n=1}^{N_l} \mathbb{1}_{\{\zeta_i\}}(\zeta_{i,n})}{N_l}.\end{aligned} \tag{42}$$

Since marginal costs are i.i.d. draws from $\psi_i(\cdot;\lambda_{i,0})$, by the strong law of large numbers,

$$\lim_{l\to\infty} \frac{\sum_{n=1}^{N_l} \mathbb{1}_{\{\zeta_i\}}(\zeta_{i,n})}{N_l} = \psi_i(\zeta_i;\lambda_{i,0}) \quad \text{a.s.} \tag{43}$$

From equations (42) and (43), we have

$$\sum_{\rho_i} \tilde{\nu}_i(\rho_i,\zeta_i)\nu_i(\rho_i,\zeta_i,\overline{\rho}_{-i}) = \psi_i(\zeta_i;\lambda_{i,0}) \quad \text{a.s.} \tag{44}$$

We conclude from equation (44) that $\tilde{\nu}_i$ is almost surely consistent for $\mathcal{G}_i(\lambda_{i,0})$.

Coarse obedience. Pick any $\zeta_i \in T$. Note that

$$\begin{aligned}&V_N(i,\zeta_i,\rho_i') - U_N(i,\zeta_i) \\ &\quad = \frac{1}{N}\sum_{n=1}^{N} \left[s(\rho_i',\overline{\rho}_{-i,n})(\rho_i' - \zeta_{i,n}) - s(\rho_{i,n},\overline{\rho}_{-i,n})(\rho_{i,n} - \zeta_{i,n})\right]\mathbb{1}_{\{\zeta_i\}}(\zeta_{i,n}) \\ &\quad = \sum_{\rho_i,\overline{\rho}_{-i}} \left[s(\rho_i',\overline{\rho}_{-i})(\rho_i' - \zeta_i) - s(\rho_i,\overline{\rho}_{-i})(\rho_i - \zeta_i)\right] Z_{i,N}(\rho_i,\zeta_i,\overline{\rho}_{-i}).\end{aligned} \tag{45}$$

Since $(\rho_{i,n},\zeta_{i,n},\overline{\rho}_{-i,n})_{n=1}^{\infty}$ has ANR, for all $\rho_i' \in A$, we have

$$\limsup_{N\to\infty} \left[V_N(i,\zeta_i,\rho_i') - U_N(i,\zeta_i)\right] \leq 0 \quad \text{a.s.} \tag{46}$$

Then, by equation (45) and inequality (46), for all $\rho_i' \in A$, we have

$$\limsup_{N\to\infty} \sum_{\rho_i,\overline{\rho}_{-i}} \left[s(\rho_i',\overline{\rho}_{-i})(\rho_i' - \zeta_i) - s(\rho_i,\overline{\rho}_{-i})(\rho_i - \zeta_i)\right] Z_{i,N}(\rho_i,\zeta_i,\overline{\rho}_{-i}) \leq 0 \quad \text{a.s.} \tag{47}$$

Moreover, on the subsequence $(Z_{i,N_l})_{l=1}^{\infty}$, for all $\rho_i' \in A$, we have

$$\begin{aligned}
&\lim_{l\to\infty} \sum_{\rho_i,\overline{\rho}_{-i}} \big[s(\rho_i',\overline{\rho}_{-i})(\rho_i'-\zeta_i) - s(\rho_i,\overline{\rho}_{-i})(\rho_i-\zeta_i)\big] Z_{i,N_l}(\rho_i,\zeta_i,\overline{\rho}_{-i}) \\
&= \sum_{\rho_i,\overline{\rho}_{-i}} \big[s(\rho_i',\overline{\rho}_{-i})(\rho_i'-\zeta_i) - s(\rho_i,\overline{\rho}_{-i})(\rho_i-\zeta_i)\big] \nu_i(\rho_i,\zeta_i,\overline{\rho}_{-i}) \qquad (48)\\
&= \sum_{\rho_i} \left[\sum_{\overline{\rho}_{-i}} \big[s_i(\rho_i',\overline{\rho}_{-i})(\rho_i'-\zeta_i) - s_i(\rho_i,\overline{\rho}_{-i})(\rho_i-\zeta_i)\big] \psi(\overline{\rho}_{-i}) \right] \tilde{\nu}_i(\rho_i,\zeta_i).
\end{aligned}$$

By inequality (47) and equation (48), for all $\rho_i' \in A$, we have

$$\sum_{\rho_i} \left[\sum_{\overline{\rho}_{-i}} \big[s_i(\rho_i',\overline{\rho}_{-i})(\rho_i'-\zeta_i) - s_i(\rho_i,\overline{\rho}_{-i})(\rho_i-\zeta_i)\big] \psi(\overline{\rho}_{-i}) \right] \tilde{\nu}_i(\rho_i,\zeta_i) \leq 0 \quad \text{a.s.} \qquad (49)$$

We conclude from equation (49) that $\tilde{\nu}_i$ is almost surely coarsely obedient for $\mathcal{G}_i(\lambda_{i,0})$. ■

F.6.2 Proof of Proposition 2

For all $\lambda_i \in \Lambda$ and $\underline{\varepsilon} := (\varepsilon(i,\zeta_i))_{\zeta_i \in T}$, where $\varepsilon(i,\zeta_i) \geq 0$ for all $\zeta_i \in T$, we have

$$\lambda_i \in \widehat{\Lambda}_{i,N}(\underline{\varepsilon}) \iff q_{i,N} \in Q_i(\underline{\varepsilon};\lambda_i) \iff R_N(i,\zeta_i) \leq \varepsilon(i,\zeta_i) \text{ for all } \zeta_i \in T. \qquad (50)$$

Pick any $\lambda \in \Lambda^*$. Under the assumptions of the proposition, for all $\zeta_i \in T$, we have (see, e.g., Section 1.2 in Faure et al. (2015)),

$$\mathbb{E}[R_N(i,\zeta_i)] \leq \frac{\Phi(i,\zeta_i)}{N}. \qquad (51)$$

By inequality (51) and Markov's inequality, for all $\varepsilon > 0$, we have

$$\mathbb{P}(R_N(i,\zeta_i) \leq \varepsilon) \geq 1 - \frac{\Phi(i,\zeta_i)}{\varepsilon N}. \qquad (52)$$

By inequality (52), for all $\overline{\alpha} \in (0,1)$,

$$\mathbb{P}(R_N(i,\zeta_i) \leq \varepsilon(i,\zeta_i)) \geq 1 - \overline{\alpha} \iff \varepsilon(i,\zeta_i) = \frac{\Phi(i,\zeta_i)}{\overline{\alpha} N}. \qquad (53)$$

For the choice of $\varepsilon(i,\zeta_i)$ on the right-hand side of equivalence (53), we have

$$\mathbb{P}\Big(R_N(i,\zeta_i) \leq \varepsilon(i,\zeta_i) \text{ for all } \zeta_i \in T\Big) = \Pi_{\zeta_i \in T} \mathbb{P}\Big(R_N(i,\zeta_i) \leq \varepsilon(i,\zeta_i)\Big) \geq (1-\overline{\alpha})^{|T|}, \qquad (54)$$

where: the equality holds because, under the assumptions of the proposition, the regrets $R_N(i,\zeta_i)$ are independent across ζ_i; the inequality follows from condition (53).

Therefore, for all $\alpha \in (0,1)$, we have

$$\mathbb{P}\Big(\lambda_i \in \widehat{\Lambda}_{i,N}(\underline{\varepsilon})\Big) \geq 1 - \alpha \iff \mathbb{P}\Big(R_N(i,\zeta_i) \leq \varepsilon(i,\zeta_i) \text{ for all } \zeta_i \in T\Big) \geq 1 - \alpha$$

$$\Longleftarrow (1-\overline{\alpha})^{|T|} = 1-\alpha$$
$$\Longleftrightarrow \overline{\alpha} = 1-(1-\alpha)^{\frac{1}{|T|}}$$
$$\Longleftrightarrow \varepsilon(i,\zeta_i) = \frac{\Phi(i,\zeta_i)}{\left[1-(1-\alpha)^{\frac{1}{|T|}}\right]N},$$

where: the first equivalence follows from the equivalences in (50); the implication follows from the chain of equalities and inequalities (54); the second equivalence follows from the choice of $\varepsilon(i,\zeta_i)$ on the right-hand side of equivalence (53). The desired result follows. ■

F.7 Alternative Assumptions on the Empirical Model

In this section, we consider an alternative empirical model in which we relax the assumptions that $\overline{\rho}_{-i,n}$ are i.i.d., that $\zeta_{i,n}$ and $\overline{\rho}_{-i,n}$ are independent, and of full feedback. Next, we present estimation results under such alternative assumptions.

F.7.1 Alternative Empirical Model

We present the empirical model and illustrate formal results and technical details by considering a typical seller i.

Regret and Asymptotic No Regret. The definitions of factual payoffs, counterfactual payoffs, regret, and asymptotic no regret are identical to those under the main specification of the empirical model in Section 5.2 (see Appendix F.6 for formal the definitions).

Bayes Coarse Correlated Equilibrium. A BCCE of the empirical stage game $\mathcal{G}_i(\lambda_i)$ is defined as follows.

Definition 17. *The probability distribution $\nu_i \in \Delta(A\times T\times\Theta)$ is a Bayes coarse correlated equilibrium of the empirical stage game $\mathcal{G}_i(\lambda_i)$ if:*

1. *ν_i is consistent for $\mathcal{G}(\lambda_i)$; that is, for all $\zeta_i \in T$, we have $\sum_{\rho_i,\overline{\rho}_{-i}} \nu_i(\rho_i,\zeta_i,\overline{\rho}_{-i}) = \psi_i(\zeta_i;\lambda_{i,0})$.*
2. *ν_i is coarsely obedient for $\mathcal{G}(\lambda_i)$; that is, for all $\zeta_i \in T$, we have*

$$\sum_{\rho_i,\overline{\rho}_{-i}} \left[s_i(\rho_i',\overline{\rho}_{-i})(\rho_i'-\zeta_i) - s_i(\rho_i,\overline{\rho}_{-i})(\rho_i-\zeta_i)\right]\nu_i(\rho_i,\zeta_i,\overline{\rho}_{-i}) \le 0 \quad \text{for all } \rho_i' \in A.$$

We denote by $E_i(\lambda_i)$ the set of BCCEs of $\mathcal{G}_i(\lambda_i)$.

ANR Pricing and Convergence. The following proposition is the equivalent of (one of the two implications in) Theorem 1 in the context of our alternative empirical model.

Proposition 3. *Suppose the sequence of prices, marginal costs, and states* $(\rho_{i,n}, \zeta_{i,n}, \overline{\rho}_{-i,n})_{n=1}^{\infty}$ *from* $\mathcal{G}_i^{\infty}$ *has ANR. Then, the sequence of empirical distributions of prices, marginal costs, and states* $(Z_{i,N})_{N=1}^{\infty}$ *converges almost surely to* $E_i(\lambda_{i,0})$.

Proof. Consider any subsequence $(Z_{i,N_l})_{l=1}^{\infty}$ of $(Z_{i,N})_{N=1}^{\infty}$ that converges to some $\nu_i \in \Delta(A \times T \times \Theta)$. We need to show that ν_i is almost surely consistent and coarsely obedient for $\mathcal{G}_i(\lambda_{i,0})$ according to Definition 17.

Consistency. Pick any $\zeta_i \in T$. Note that

$$\sum_{\rho_i, \overline{\rho}_{-i}} \nu_i(\rho_i, \zeta_i, \overline{\rho}_{-i}) = \sum_{\rho_i, \overline{\rho}_{-i}} \lim_{l \to \infty} Z_{i,N_l}(\rho_i, \zeta_i, \overline{\rho}_{-i}) = \lim_{l \to \infty} \frac{\sum_{n=1}^{N_l} \mathbb{1}_{\{\zeta_i\}}(\zeta_{i,n})}{N_l}.$$

That ν_i is almost surely consistent for $\mathcal{G}_i(\lambda_{i,0})$ follows from the same argument as that in the proof of consistency in Proposition 1 (see Appendix F.6).

Coarse obedience. Pick any $\zeta_i \in T$. By the same steps as those in the proof of coarse obedience in Proposition 1 (see Appendix F.6), for all $\rho_i' \in A$, we have

$$\limsup_{N \to \infty} \sum_{\rho_i, \overline{\rho}_{-i}} \left[s(\rho_i', \overline{\rho}_{-i})(\rho_i' - \zeta_i) - s(\rho_i, \overline{\rho}_{-i})(\rho_i - \zeta_i)\right] Z_{i,N}(\rho_i, \zeta_i, \overline{\rho}_{-i}) \leq 0 \quad \text{a.s.} \tag{55}$$

Moreover, on the subsequence $(Z_{i,N_l})_{l=1}^{\infty}$, for all $\rho_i' \in A$, we have

$$\begin{aligned} &\lim_{l \to \infty} \sum_{\rho_i, \overline{\rho}_{-i}} \left[s(\rho_i', \overline{\rho}_{-i})(\rho_i' - \zeta_i) - s(\rho_i, \overline{\rho}_{-i})(\rho_i - \zeta_i)\right] Z_{i,N_l}(\rho_i, \zeta_i, \overline{\rho}_{-i}) \\ &= \sum_{\rho_i, \overline{\rho}_{-i}} \left[s(\rho_i', \overline{\rho}_{-i})(\rho_i' - \zeta_i) - s(\rho_i, \overline{\rho}_{-i})(\rho_i - \zeta_i)\right] \nu_i(\rho_i, \zeta_i, \overline{\rho}_{-i}). \end{aligned} \tag{56}$$

By inequality (55) and equation (56), for all $\rho_i' \in A$, we have

$$\sum_{\rho_i, \overline{\rho}_{-i}} \left[s(\rho_i', \overline{\rho}_{-i})(\rho_i' - \zeta_i) - s(\rho_i, \overline{\rho}_{-i})(\rho_i - \zeta_i)\right] \nu_i(\rho_i, \zeta_i, \overline{\rho}_{-i}) \leq 0 \quad \text{a.s.} \tag{57}$$

We conclude from equation (57) that ν_i is almost surely coarsely obedient for $\mathcal{G}_i(\lambda_{i,0})$. ■

Estimation, Confidence Region, and Coverage. The following proposition, which is the equivalent of Theorem 3 in the context of our alternative empirical model, characterizes a uniformly valid confidence region for all $\lambda_i \in \Lambda_i^*$.

Proposition 4. *In the alternative empirical model, suppose: 1. adversarial bandit environment; 2. bandit feedback; 3. the algorithm optimizes the convergence rate of the expected regret to 0 given 1. and 2. Fix* $\alpha \in (0,1)$ *and* $N \in \mathbb{N}$. *For all* $\zeta_i \in T$ *and* $\lambda_i \in \Lambda$, *let*

$$\varepsilon_{\alpha,N}(\zeta_i; \lambda_i) := \frac{\left[K(i, \zeta_i)\sqrt{\ln |A|}\sqrt{\psi(\zeta_i; \lambda_i)}\right] |T|}{\alpha \sqrt{N}}.$$

Then, $\widehat{\Lambda}_{i,N}(\underline{\boldsymbol{\varepsilon}}_{\alpha,N}) := \left\{\lambda_i \in \Lambda : q_{i,N} \in Q_i(\underline{\boldsymbol{\varepsilon}}_{\alpha,N}(\lambda_i); \lambda_i)\right\}$ *is a uniformly valid confidence region for all* $\lambda_i \in \Lambda^*$*; that is,*

$$\inf_{\lambda \in \Lambda^*} \mathbb{P}\Big(\lambda_i \in \widehat{\Lambda}_{i,N}(\underline{\boldsymbol{\varepsilon}}_{\alpha,N})\Big) \geq 1 - \alpha.$$

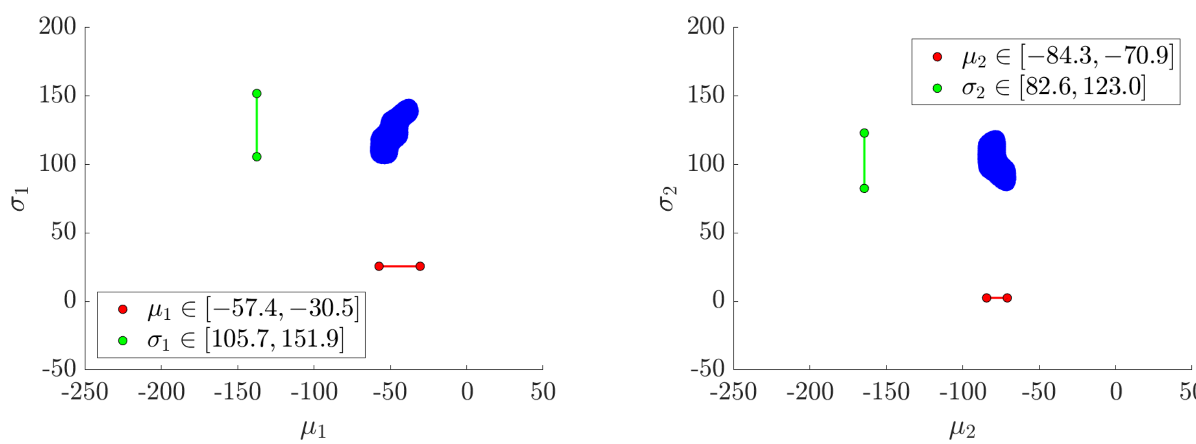
Proof. The proof mimics that of Theorem 3 in Appendix A.3 ■

The proposition shows that the different assumptions adopted in this appendix result in a convergence rate of $\mathcal{O}\Big(\frac{1}{\sqrt{N}}\Big)$, as opposed to $\mathcal{O}\Big(\frac{1}{N}\Big)$ for the assumptions in our preferred specification. Therefore, the implied worst-case bounds on regrets are unlikely to yield informative inferences for the sample sizes in our application. To address this, we rely on our simulation results in Section 4.5, where we found that average empirical regrets are only a fraction of the theoretical worst-case regret. We compute this ratio r in our simulations, and apply it to the $\varepsilon_{\alpha,N}(\zeta_i; \lambda_i)$, thus using $r\varepsilon_{\alpha,N}(\zeta_i; \lambda_i)$ to compute our confidence regions.

F.7.2 Estimation Results under the Alternative Empirical Model

In panels (a) and (b) of Figure 13, we represent the confidence regions for the top-2 Swappa sellers in our data under the alternative assumptions presented in this appendix. For each seller i, the figure plots (in blue) the confidence region $\widehat{\Lambda}_{i,N}(\underline{\boldsymbol{\varepsilon}}_{\alpha,N})$ in the (μ_i, σ_i)-space, the parameters characterizing the seller-specific distribution of marginal costs. We also represent (in red and green) the projections of the confidence regions for each of the two parameters. The estimated confidence regions are qualitatively similar to those in Figure 7.

Figure 13: Confidence Regions.



(a) Seller 1 (b) Seller 2

We plot in blue confidence regions in the (μ_i, σ_i)-space; we report in red and green projections for μ_i and σ_i. Each panel corresponds to one of the top-2 Swappa sellers. Computational details are in Appendix E.3.

We also compute the distribution of marginal cost residuals and marginal costs corresponding to the confidence regions, similarly to Table 1. The results, reported in Table 8, are again similar to those under our preferred specification in the main text.

Table 8: Estimated Distributions of Marginal Cost Residuals and Marginal Costs.

	Marginal Cost Residuals ($)		Marginal Cost ($)	
	Mean	SD	Mean	SD
	(1)	(2)	(3)	(4)
Seller 1	$[-91.3, -65.8]$	$[88.3, 113.3]$	$[504.6, 530.1]$	$[206.4, 226.6]$
Seller 2	$[-102.8, -72.5]$	$[76.2, 107.7]$	$[424.2, 454.5]$	$[257.3, 278.7]$

We report statistics of marginal cost residuals (columns 1-2), and marginal costs (columns 3-4) for the parameter estimates corresponding to the assumptions in this appendix. Each of the top-2 Swappa sellers in our data corresponds to a row. Marginal cost residual statistics are obtained as in Table 1.

Appendix References

Cesa-Bianchi, N. and G. Lugosi (2006): *Prediction, Learning, and Games*, Cambridge University Press.

Faure, M., P. Gaillard, B. Gaujal, and V. Perchet (2015): "Online Learning and Game Theory. A Quick Overview with Recent Results and Applications," *ESAIM: Proceedings and Surveys*, 51, 246–271.

Grant, M. and S. Boyd (2008): "Graph Implementations for Nonsmooth Convex Programs," in *Recent Advances in Learning and Control*, ed. by V. Blondel, S. Boyd, and H. Kimura, Springer-Verlag Limited, Lecture Notes in Control and Information Sciences, 95–110, http://stanford.edu/~boyd/graph_dcp.html.

——— (2014): "CVX: Matlab Software for Disciplined Convex Programming, version 2.1," https://cvxr.com/cvx.

Kvasnica, M., P. Grieder, and M. Baotić (2004): "Multi-Parametric Toolbox (MPT)," .

Nisan, N., T. Roughgarden, E. Tardos, and V. V. Vazirani (2007): *Algorithmic Game Theory*, Cambridge University Press.

Roughgarden, T. (2016): *Twenty Lectures on Algorithmic Game Theory*, Cambridge University Press.