Empirical Framework for Two-Player Repeated Games with Random States

Arkadiusz Szydłowski*

University of Leicester

October 26, 2022

Abstract

We provide methods for the empirical analysis of a class of two-player repeated games with i.i.d. shocks, allowing for non-Markovian strategies. The number of possible equilibria in these games is large and, usually, theory is silent about which equilibrium will be chosen in practice. Thus, our method remains agnostic about selection among these multiple equilibria, which leads to partial identification of the parameters of the game. We propose a profiled likelihood criterion for building confidence sets for the structural parameters of the game and derive an easily computable upper bound on the critical value. We demonstrate good finite-sample performance of our procedure using a simulation study. We illustrate the usefulness of our method by studying the effect of repealing the Wright Amendment on entry and exit into Dallas airline markets and find that the static game approach overestimates the negative effect of the law on entry into these markets.

Keywords: Empirical games, Partial identification, Parameter on boundary, Airline market

* *E-mail*: ams102@le.ac.uk, *Address*: Department of Economics, Finance and Accounting, University of Leicester, University Road, Leicester, LE1 7RH, UK

1 Introduction

The theory of repeated strategic interactions has been thoroughly studied by economists. However, the empirical analysis of repeated games is scarce. Repeated games usually feature many equilibria making their analysis difficult. In this paper we endeavour to provide an empirical framework for a two-player binary repeated game with i.i.d. shocks to payoffs. We leverage the theoretical result that, conditional on the continuation payoff, a repeated game reduces to a form of a static game for identification, develop a convenient maximum likelihood characterization of the identified set and use it to develop a profiled likelihood ratio test for estimating confidence sets for the structural parameters. We derive the asymptotic distribution of the test statistic and provide an upper bound on the critical value. Monte Carlo simulations suggest that this bound leads to only moderately conservative inference. However, this critical value requires a preliminary estimator of the identified set which makes it computationally intensive. Thus, we also suggest an alternative easily computable, but more conservative, upper bound on the critical value.

As a static game Nash equilibrium repeated over time constitutes an equilibrium in the repeated version of such game, allowing repeated interactions between agents enlarges the set of possible equilibria and allows the model to justify wider range of behaviours. Since repeated games are notoriously perceived as providing very non-sharp theoretical predictions, one may think that in fact we may be able able to justify any behaviour with such model. However, play observed in the data may exclude some equilibria. Take as an example a classic repeated prisoners dilemma with no stochastic shocks (or very small shocks). If we observe in the data that players cooperate in this game, this rules out a unique static equilibrium in this game, namely playing defect-defect. Thus, combining theory and data may still lead to useful predictions. This example also illustrates that predictions from a static and repeated version of a game can substantially differ. In fact, in Section 8 we re-investigate the policy analysis in Ciliberto & Tamer (2009) of the effect of repealing the Wright Amendment on entry and exit into Dallas airline markets by Southwest and American using our empirical repeated game and find that the static game approach overestimates the negative effect of the law on entry into these markets.[1]

Our article is related to the literature on estimation of dynamic games with incomplete informa-

[1] Recent papers analysing the US airline market include Goolsbee & Syverson (2008), Gerardi & Shapiro (2009), Berry & Jia (2010) and Ciliberto et al. (2021).

tion. However, it differs from the approach in Bajari et al. (2007) in important aspects. We focus on a simple form of a dynamic game without state dependence and with complete information.[2] In this subset of dynamic games we do not have to restrict ourselves to Markov perfect equilibria, allowing for a rich set of equilibrium strategies potentially dependent on the whole history of play. In fact, Markov perfect equilibria in this game correspond to stage game equilibria, thus Markov restriction makes the dynamic dimension irrelevant.

Using experimental setup and an estimated structural model Salz & Vespa (2020) show that the Markovian assumption is often not satisfied and may lead to erroneous estimates and predictions, which motivates considering models without this assumption. Our model delivers a framework allowing non-Markov play, which, for example, facilitates justification of collusive behavior. We stress however that we restrict ourselves to stationary-outcome equilibria. Thus, our approach can be seen as an alternative to the established dynamic games approach in settings where agents face a long-term strategic interaction in a stable environment with well known structure of the payoffs.

Our inference approach is similar to Kline & Tamer (2016) and Chen et al. (2018) in that we assume that the (population) likelihood identifies the choice probabilities but not the parameters (see also Giacomini & Kitagawa (2021)). We apply the same profiled likelihood ratio criterion as in Chen et al. (2018) but analyse it differently with a view of obtaining a practical inference method tailored to our repeated game model, as the general procedure in the latter paper is computationally difficult in our setup. The computational advantage of our method comes at the cost of conservative inference, though. This is a trade-off frequently encountered in applications of partially identified models.

Since we focus on providing marginal confidence intervals for the identified structural parameters, our article is related to the literature on marginal inference in partially identified models – see e.g. Bugni et al. (2017) and references therein. Our problem features also likelihood ratio statistic with a parameter on the boundary of the parameter space as in e.g. Shapiro (1985), Andrews (2001) (in point identified model) and in e.g. Chernozhukov et al. (2007) (in partially identified model). Recent contribution to this literature involves e.g. Al Mohamad et al. (2020). The technique in the latter paper could, in principle, be used to bound the critical value for our problem

[2]In principle, our method can be extended to more general dynamic games with non-trivial state dependence, however implementing such an extension poses a difficult computational challenge so we leave it for further research.

as well. However, their method would work well in settings when only a few boundary constraints are binding, which makes it unattractive in our setup.

An alternative approach to inference in our model would characterize the identified set using moment inequalities like in Beresteanu et al. (2011) or Galichon & Henry (2011) and then employ profiled inference methods (Romano & Shaikh (2008), Bugni et al. (2017) or Kaido et al. (2019)).

Other articles addressing identification in repeated games include Abito & Chen (2021) and Lee & Stewart (2016). The former work effectively analyses a non-stochastic repeated game and provides worst-case bounds for probabilities of play across heterogenous repeated games. We allow payoffs to be stochastic, even if with limited dependence, as the assumption of time-invariant, deterministic payoffs seems too strong for applications. We also provide a sharp characterization of the identified set, rather than worst-case bounds. Lee & Stewart (2016), again, consider a non-stochastic repeated game and, additionally, assume that players' best response correspondences are fully observed, which is rarely the case in empirical applications.

2 Repeated Game with Random States

A defining feature of our model is that agents, unlike in a standard repeated game, play a different game every period. The payoffs in the stage game are stochastic and distributed i.i.d. over time (conditional on observables). Payoffs are fully observed by the players but the econometrician observes only the history of play and has only limited information about the non-stochastic part of the payoff. In particular, she does not observe the realization of the i.i.d. shocks.[3]

We focus on binary games. Let $\mathcal{A} = \{0,1\}^2$ denote the set of actions with a typical element a. There are two players identified by $i = 1, 2$ playing an infinitely repeated game, $t = 0, 1, 2 \ldots, \infty$. Every period a random payoff relevant vector $\varepsilon_t = \{(\varepsilon_{1t,a}, \varepsilon_{2t,a})\}_{a \in \mathcal{A}}$ is drawn from the distribution $F_\varepsilon : \mathcal{E} \to [0, 1]$ with the following properties:

Assumption ID1. *ε_t are i.i.d. across time and $E(\varepsilon_{t,a}) = \mathbf{0}$ for all a and $\varepsilon_{t,\mathbf{0}} = \mathbf{0}$ (normalization).*

Denote player i's payoff in period t by $\tilde{u}_i(a_t, \varepsilon_{it,a_t}; \alpha_i) = u_i(a_t; \alpha_i) - \varepsilon_{it,a_t}$, where α_i is a finite dimensional parameter. Define $\alpha = (\alpha_1, \alpha_2)$. In practice some observed characteristics $X \in \mathcal{X} \subset R^K$ will usually enter the payoff function. For now we suppress them from notation for the ease

[3]For the rest of the paper we will simply refer to our repeated game with random states as a "repeated game".

of exposition. We discuss the additional issues related to the presence of covariates in Section 6. Payoffs in future periods are discounted by a common discount factor $\delta \in (0,1)$.

Let $\mathcal{H}^t = (\mathcal{E} \times \mathcal{A})^t$ denote the set of period t ex ante histories with a typical element $h^t = \{\varepsilon_{s,a_s}, a_s\}_{s=0}^{t-1}$. Let $\tilde{\mathcal{H}}^t = (\mathcal{E} \times \mathcal{A})^t \times \mathcal{E}$ denote the set of period t ex post histories with a typical element $\tilde{h}^t = (h^t, \varepsilon_t)$. A pure strategy profile, σ, is a pair of mappings from $\tilde{\mathcal{H}}^t$ to $\mathcal{A}$. Player i's expected lifetime payoff from playing the game is:

$$U_i(\sigma; \alpha_i) = E^{\sigma}\left[\sum_{t=0}^{\infty} \delta^t \tilde{u}_i(a_t, \varepsilon_{it,a_t}; \alpha_i)\right]$$

where the expectation is taken over the histories induced by σ.

A normalized continuation payoff of player i after a history $\tilde{h}^t$ is given by:

$$V_i^{\sigma}(\tilde{h}^t; \alpha_i) = (1-\delta)u_i(\sigma(\tilde{h}^t); \alpha_i) - (1-\delta)\varepsilon_{it,\sigma(\tilde{h}^t)} + \delta \int V_i^{\sigma}(\tilde{h}^{t+1}; \alpha_i) dF_{\varepsilon}$$

In our game with i.i.d. shocks the expected continuation payoff of player i, namely $\int V_i^{\sigma}(\tilde{h}^{t+1}; \alpha_i) dF_{\varepsilon}$ in the above expression, depends only on a current action $a = \sigma(\tilde{h}^t)$ so we can denote it by $v_i(a)$ (see Remark 5.7.1 in Mailath & Samuelson (2006)).

3 Identification

We start with an example to discuss the main ideas behind identification and formalize it later.

3.1 Illustrative example

Consider a simple game with the following stage game payoffs (for further reference, we will call this game Σ^S):

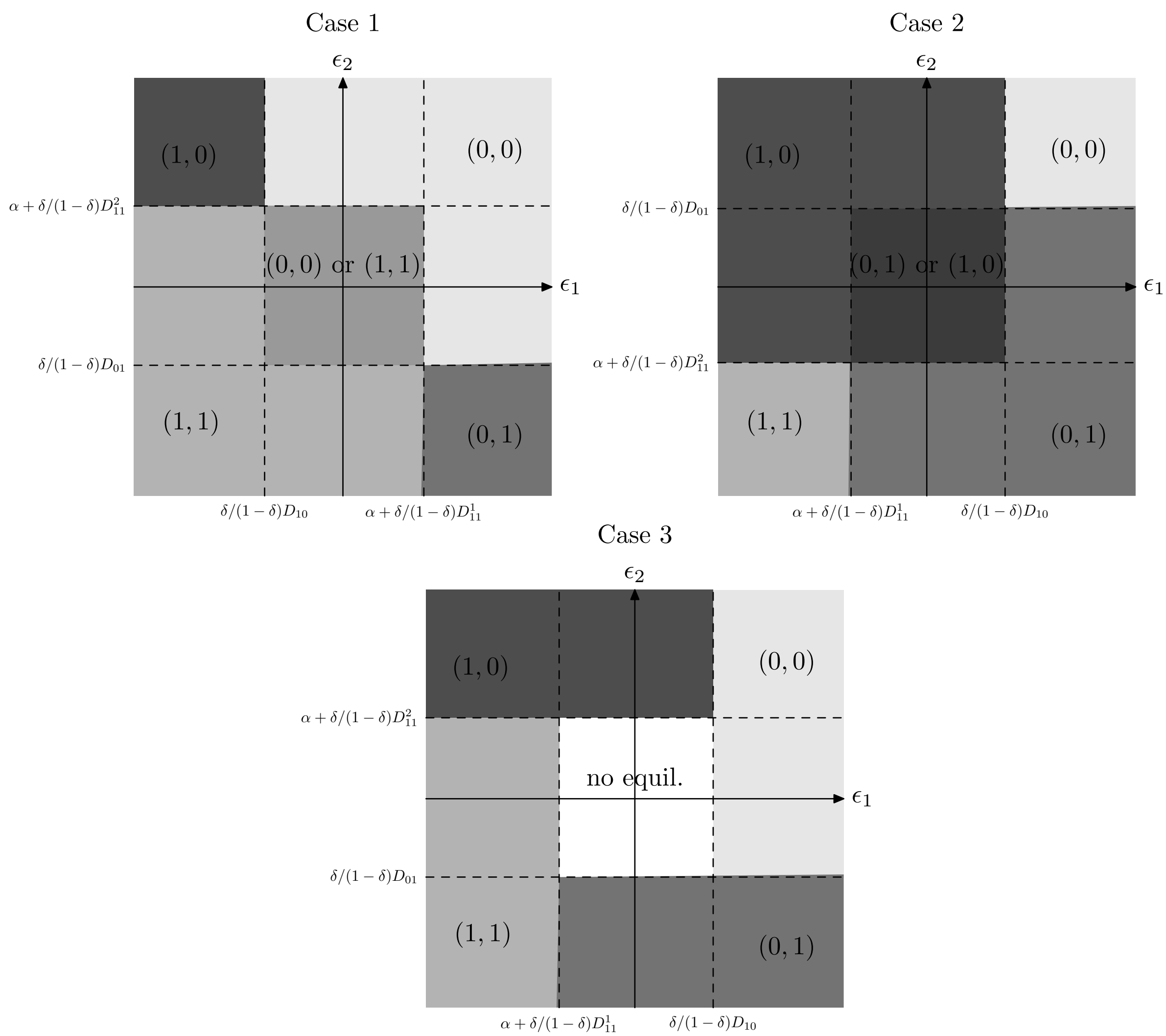
		P 2: 1	P 2: 0
P 1	1	$\alpha_1 - \varepsilon_1, \alpha_2 - \varepsilon_2$	$-\varepsilon_1, 0$
	0	$0, -\varepsilon_2$	$0, 0$

where ε_1 and ε_2 are identically and independently distributed (with some abuse of notation, let F_ε denote their distribution). The corresponding normal form of the repeated game is given by:

		P 2: 1	P 2: 0
P 1	1	$(1-\delta)(\alpha_1-\varepsilon_1)+\delta v_1(1,1)$, $(1-\delta)(\alpha_2-\varepsilon_2)+\delta v_2(1,1)$	$-(1-\delta)\varepsilon_1+\delta v_1(1,0)$, $\delta v_2(1,0)$
	0	$\delta v_1(0,1)$, $-(1-\delta)\varepsilon_2+\delta v_2(0,1)$	$\delta v_1(0,0), \delta v_2(0,0)$

Let $\alpha_1 = \alpha_2 = \alpha > 0$. By analogy to static binary games (see Tamer (2003)), we may have multiple Nash equilibria in this normal form. For example, assuming that players play a static Nash equilibrium in every period, i.e. $v_i(a) = 0$ for all i and a, both (1,1) and (0,0) can arise in equilibrium when ε_1 and ε_2 are between 0 and α.

Figure 1: Multiple equilibria in the normal form



In general, the region of ε for which we will have multiple equilibria depends on V, or more specifically – on the value contrasts D_a^i:

$$D_{10} = v_1(1,0) - v_1(0,0), \quad D_{01} = v_2(0,1) - v_2(0,0)$$
$$D_{11}^1 = v_1(1,1) - v_1(0,1), \quad D_{11}^2 = v_2(1,1) - v_2(1,0)$$

Figure 1 shows regions of values of $(\varepsilon_1, \varepsilon_2)$ for which different equilibria occur for different values of D_a^i's. There are several regions where the game has two equilibria. Let us focus on Case 1 and the probability that (1,1) is played in a subgame-perfect equilibrium. For simplicity assume that only symmetric equilibria are played in this game so $D_{11}^1 = D_{11}^2 = D_{11}, D_{10} = D_{01}$. Let π denote the probability that (0,0) is played in equilibrium in the region where both (0,0) and (1,1) can be played. Then, we can write the probability of (1,1) being played in this repeated game, conditional on the continuation values V as:

$$p(1,1|V) = \left[F_\varepsilon\left(\alpha + \frac{\delta}{1-\delta}D_{11}\right)\right]^2 - \pi\left[F_\varepsilon\left(\alpha + \frac{\delta}{1-\delta}D_{11}\right) - F_\varepsilon\left(\frac{\delta}{1-\delta}D_{01}\right)\right]^2$$

Thus, conditional on V the problem boils down to a standard identification problem in static games, where we may have multiple equilibria and we do not observe the equilibrium selection probability π. However, in our repeated game lack of identification is magnified by the fact that we observe neither π nor V. Nevertheless, our application shows that this does not preclude us from obtaining meaningful bounds on the model parameters.

3.2 General case

Let us now formalize the identification argument. Throughout our analysis we will assume that

Assumption ID2. ***(a)*** *δ and F_ε are known,* ***(b)*** *$u(a_i, a_{-i}; \alpha) = 0$ for some $a_i \in \mathcal{A}$ and all $a_{-i} \in \mathcal{A}$*

This assumption is frequently made in dynamic discrete choice and dynamic games literature. In the case of dynamic discrete choice models, Magnac & Thesmar (2002) showed that the payoff functions are not identified without knowledge of the discount factor and distribution of shocks or without normalizing the payoffs for one of the alternatives. In practice one can conduct the analysis for different values of δ and different distributions F_ε in order to investigate the sensitivity of results to this assumption.

We assume that we observe a finite history of play, $\{a_s\}_{s=0}^{t-1}$, in at least one game (market) and the observed play is generated by some subgame-perfect Nash equilibrium. In principle, characterizing the set of equilibria in our game is complicated since there are numerous strategies that can generate a given equilibrium outcome. Fortunately, we do not have to work with **strategies** directly but rather characterize sufficient and necessary condition(s) for observed **choices** to be made in equilibrium. Let $p_t(a)$ denote the probability of a being played in equilibrium at some t. We will restrict this probability to be stationary. Note that this does not necessarily restrict strategies to be stationary and allows, for example, grim-trigger strategies (see e.g. Example 2.4.1 in Mailath & Samuelson (2006)), as long as they lead to a stationary outcome on the equilibrium path.

Assumption ID3. *Players play a stationary-outcome subgame-perfect Nash equilibrium, i.e.* $p_t(a) = p(a)$ *for all* t*.*

In principle, we could work with non-stationary outcome equilibria. However, this would require that we observe the same game being played in many different markets such that we can calculate $p_t(a)$ for every t by looking at probability of a being played across markets. Such data is hard to come by since payoffs usually differ between markets due to different market characteristics. In some games Assumption ID3 rules out some important equilibria, e.g. some efficient symmetric equilibria (see Section 6.3 in Mailath & Samuelson (2006)), thus one needs to be careful if it is not too strong in a particular game of interest. Generally, it is difficult to characterize strategies excluded by this assumption.

Before we characterize the empirical content of our model, note that, since the researcher observes neither present nor historical ε's, it is not possible to restrict the set of potential equilibria based on the observed history of play, i.e. based on observing history of actions only. In other words, a given observed history of actions $\{a_s\}_{s=1}^{t}$ can be reconciled with any $V \in \mathcal{V}$ by choosing the history of ϵ's appropriately (e.g. driving some close to $\pm\infty$). Thus, all the potential equilibrium continuation payoffs have to be considered at each time period and history.

Lemma 1. *Let* $\mathcal{V}_S(\alpha)$ *denote the set of pairs of expected lifetime payoffs that can be reached in a stationary-outcome equilibrium, i.e. for any subgame-perfect Nash equilibrium strategy* σ *we have* $(U_1(\sigma;\alpha_1), U_2(\sigma;\alpha_2)) \in \mathcal{V}_S(\alpha)$*, and* $v(a) = (v_1(a), v_2(a)) \in \mathcal{V}_S(\alpha)$ *denote expected continuation pay-*

offs associated with playing a in the current period. Define $V = (v(a), v(a_1', a_2), v(a_1, a_2'), v(a_1', a_2'))$ with $a_1 \neq a_1', a_2 \neq a_2'$. Then, for some distribution $F_V : \mathcal{V}_S(\alpha)^4 \to [0,1]$:

$$p(a) = \int_{\mathcal{V}_S(\alpha)^4} P(a \text{ is a Nash equilibrium in the normal form game with payoffs } g_i(a) | V = v) dF_V(v) \tag{1}$$

where:

$$g_i(a) = (1-\delta)(u_i(a;\alpha_i) - \varepsilon_{i,a}) + \delta v_i(a),$$

and $\{(\varepsilon_{1,a}, \varepsilon_{2,a})\}_{a \in \mathcal{A}}$ are drawn from F_ε.

Lemma 1 follows from Proposition 5.7.3 in Mailath & Samuelson (2006). The distribution F_V can be interpreted as an equilibrium selection function since every $v \in \mathcal{V}_S(\alpha)$ corresponds to a different subgame-perfect equilibrium in our stochastic game. If data comes only from a single game and mixing between different equilibria is not allowed, then F_V is degenerate on some particular V. However, if we allow mixing between equilibria data (for example, across different markets), then F_V may put non-trivial mass on several V's corresponding to different markets.

Lemma 1 is useful for two reasons. First, it allows us to describe the identified set in the model without dealing with strategy functions, which belong to a complicated functional space. Instead, in order to verify if a given parameter value α can be reconciled with observed probabilities, it is enough to check if these probabilities can be generated in equilibrium by some (combination of) continuation payoffs in $\mathcal{V}_S(\alpha)^4$. Second, calculating the sets $\mathcal{V}_S(\alpha)$ is relatively easy for a game with discretely supported independent shocks and we can approximate these sets for a game with continuous F_ε by increasing the number of support points of ε. We discuss methods for obtaining these sets in the next section.

Thinking of the relationship between V in our model and an equilibrium selection probability in a classic static game model, the main difference seems to be that the domain of V depends on the parameters of the game α, whereas in the static model the domain of the selection probability is always $[0,1]$. This complicates inference as the set $\mathcal{V}$ has to be re-calculated (or approximated) for each candidate value of α.

Note that the normal form game may have multiple equilibria itself and denote by π the equilibrium selection probability in the ambiguous region. Let us now make explicit the dependence of the choice probabilities on the parameter α and the equilibrium selection probabilities π and F_V by writing $p(a; \alpha, \pi, F_V)$. $\mathcal{F}(\alpha)$ denotes the set of all cumulative distribution functions supported on $\mathcal{V}_S(\alpha)^4$. As a result of Proposition 1 and the standard argument for identification of maximum likelihood (see e.g. Amemiya (1986)), we have the following corollary:

Corollary 1. *Define the identified set Θ_{01}^S as the set of α's for which observed probabilities are consistent with some stationary-outcome Nash subgame-perfect equilibrium. Assume that:*

$$\sup_{\pi \in [0,1], F_V \in \mathcal{F}(\alpha)} E_0 |\log p(a; \alpha, \pi, F_V)| < \infty$$

for all $\alpha \in \Theta_{01}^S$ and that the selection probability π does not depend on ε or V. Then:

$$\Theta_{01}^S = \arg\sup_{\alpha} \sup_{\pi \in [0,1], F_V \in \mathcal{F}(\alpha)} E_0 \log p(a; \alpha, \pi, F_V) \tag{2}$$

where the expectation is taken with respect to the true distribution of a. This set is sharp.

It is worth mentioning that a sharp characterisation of the identified set can also be obtained through moment inequalities using the random set approach of Beresteanu et al. (2011), which does not require putting, arguably, strong restrictions on the dependence between π and (ε, V) present in Corollary 1.[4] However, with that approach it is less clear how to obtain easily computable critical values or moderately conservative approximations to these critical values as in Section 5.1. We compare our approach to the moment inequality approach in Appendix C.

4 Equilibrium continuation payoff sets

In this section we illustrate how one can obtain the sets $\mathcal{V}_S(\alpha)$. In fact, we will show how to obtain an outer approximation to this set, $\mathcal{V}(\alpha)$, since we will allow $\mathcal{V}(\alpha)$ to contain continuation payoffs both from stationary- and non-stationary-outcome equilibria. Having calculated $\mathcal{V}(\alpha)$, we can characterize an outer set of the identified set, Θ_{01}, by replacing $\mathcal{V}_S(\alpha)$ with $\mathcal{V}(\alpha)$ in characterisation

[4]One could relax this assumption and allow π to be a parametric function of V and still keep inference based on Corollary 1 computationally appealing. Allowing π to vary with ε seems more difficult.

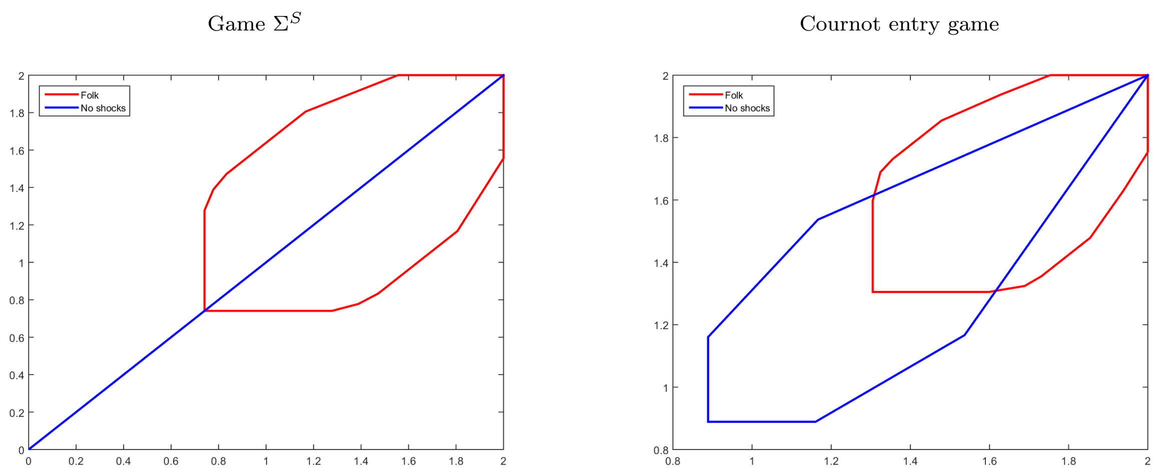
(2). Even though $\Theta_{01}^S \subset \Theta_{01}$ the outer set obtained in this way is still quite narrow in our application (and Monte Carlo simulations).

With finite discrete support of shocks we can view games with different draws of ε as separate non-stochastic repeated games and use the algorithms introduced in Abreu et al. (1990) and Abreu & Sannikov (2014) to find the sets of equilibrium payoffs in these games. Let $\mathcal{V}_{\varepsilon^m}(\alpha), m = 1, \ldots, M$ denote the collection of equilibrium payoff sets for different values of shocks $\varepsilon^m \in \mathcal{E}$. Then, the set of equilibrium payoffs in the full game can be calculated as a Minkowski sum:

$$\mathcal{V}(\alpha) = \{V : V = V_1 \cdot P(\varepsilon = \varepsilon^1) + \ldots + V_M \cdot P(\varepsilon = \varepsilon^M), V_1 \in \mathcal{V}_{\varepsilon^1}(\alpha), \ldots, V_M \in \mathcal{V}_{\varepsilon^M}(\alpha)\}$$

(see Remark 5.7.1 in Mailath & Samuelson (2006)).[5] We will sometimes refer to this set of equilibrium payoffs as the '$\mathcal{V}$ set'.

Figure 2: Set of equilibrium payoffs, $\mathcal{V}$



Note: $\alpha = 1, \delta = 0.75$. Blue line demarcates the set of payoffs in the non-stochastic version of the game with $\varepsilon_1 = \varepsilon_2 = 0$. Red line delineates the set in the stochastic game.

We will illustrate the construction of this set using the game Σ^S and a simplified version of an entry game in which entrants engage in Cournot competition (see Tamer (2003), henceforth, referred to as "Cournot entry game"), both with discrete support of shocks over $\mathcal{E} = \{-2, 0, 2\}^2$.

[5] According to Mailath & Samuelson (2006) the above equality holds for a fixed M. Showing that it also holds as $M \to \infty$ or providing a bound on the finite approximation is an exercise in the theory of repeated games, as this would involve showing that the limit of the approximation is an equilibrium continuation set in a game with continuously distributed ε, a task we leave for further research. One can always see our model as a model with discrete ε and a large number of support points.

The stage game payoffs in the latter game are given by:

		P 2	
		1	0
P 1	1	$\alpha - \varepsilon_1, \alpha - \varepsilon_2$	$\frac{9}{4}\alpha - \varepsilon_1, 0$
	0	$0, \frac{9}{4}\alpha - \varepsilon_2$	$0, 0$

Figure 2 compares the equilibrium payoff sets of the non-stochastic and stochastic version of game Σ^S (left panel) and the Cournot entry game (right panel). Note that due to symmetry the $\mathcal{V}$ set in Σ^S without shocks is just a straight line connecting stage game payoffs in the two Nash equilibria. However, once we add the stochastic shocks the game is not necessarily symmetric and the $\mathcal{V}$ set will have a non-empty interior for some values of ε. As a result, the $\mathcal{V}$ set in the stochastic game has a non-empty interior as well. The $\mathcal{V}$ sets in the stochastic game exclude some low payoff pairs from the non-stochastic version since now players can condition their strategies on the realizations of the shocks and, as a result, achieve better outcomes.

We can view a game with finite support of shocks as an approximation of a game with continuous F_ε where the approximation becomes more precise as the number of support points, M, grows. For the purpose of inference we can make $M \to \infty$ as the sample size $T \to \infty$.

Finally, it is worth noting that, although the probabilities $p(a)$ depend only on the continuation value contrasts, e.g. $D^1_{11} = v_1(1,1) - v_1(1,0), D^2_{11} = v_2(1,1) - v_2(0,1)$, working with the value contrasts as a primitive of the game is not that convenient because the restriction $V \in \mathcal{V}(\alpha)$[4] may put some non-trivial restrictions on the set of equilibrium D's. For example, setting $D^1_{11} = \max_{(v_1,v_2)\in\mathcal{V}} v_1 - \min_{(v_1,v_2)\in\mathcal{V}} v_1$ and $D^2_{11} = \max_{(v_1,v_2)\in\mathcal{V}} v_2 - \min_{(v_1,v_2)\in\mathcal{V}} v_2$ may not be feasible since the points $(\max_{(v_1,v_2)\in\mathcal{V}} v_1, \max_{(v_1,v_2)\in\mathcal{V}} v_2)$ and $(\min_{(v_1,v_2)\in\mathcal{V}} v_1, \min_{(v_1,v_2)\in\mathcal{V}} v_2)$ may not belong to the $\mathcal{V}$ set. Thus, we work with the equilibrium payoffs V rather than value contrasts because this allows us to handle the constraints more easily.

5 Inference

We suggest using maximum likelihood for inference. In practical applications some components of the payoff vector, X, are observed. They usually enter payoffs linearly and we are interested in estimating their coefficients, $\beta = (\beta_1, \beta_2)$. Thus, all the probabilities below should condition on X.

We suppress this conditioning and discuss adding covariates to the model in Section 6.

We assume that the discount factor δ is known so the goal is to estimate a confidence set for $\theta_1 \equiv (\alpha, \beta), \theta_1 \in \Theta_1 \subset \mathcal{R}^{d_1}$. We allow for any equilibrium selection mechanism $\pi \in [0,1]$ but restrict F_V to be degenerate:

Assumption INF1. *All observations are generated from a unique V.*

If the data is pooled across markets, the latter restriction imposes that the same dynamic equilibrium is played in different markets (similarly to Bajari et al. (2007)).[6] Let $\theta_2 \equiv (\pi, V) \in \Theta_2(\theta_1)$ where $\Theta_2(\theta_1)$ is the correspondence mapping values of θ_1 to the sets of corresponding continuation values V and the set of feasible selection probabilities π. Further, $\theta = (\theta_1, \theta_2) \in \Theta \subset \mathcal{R}^d, d = d_1 + d_2$. As in Chen et al. (2018) we suggest using supremum of the profiled likelihood ratio statistic for inference and building the confidence set as a collection of points for which the LR statistic falls below the critical value from the asymptotic distribution of this sup-profiled statistic.

We observe a sample of action pairs $\{Y_t\}_{t=1}^T$. Let $p(a|\theta)$ denote the model implied probabilities when parameters are set to θ. Denote $\gamma_a(\theta) = p(a|\theta) - p(a|\theta_0)$ for $\theta_0 \in \Theta_0$, where Θ_0 is the identified set for θ, and let the vector $\gamma(\theta) \equiv [\gamma_{11}(\theta) \quad \gamma_{10}(\theta) \quad \gamma_{01}(\theta)]' \in \Gamma$ collect the choice probabilities. Then the likelihood of observing a given Y_t can be written as:

$$p(Y_t, \theta) = \gamma_{11}(\theta)^{Y_{1t}Y_{2t}} \gamma_{10}(\theta)^{Y_{1t}(1-Y_{2t})} \gamma_{01}(\theta)^{(1-Y_{1t})Y_{2t}} (1 - \gamma_{11}(\theta) - \gamma_{10}(\theta) - \gamma_{01}(\theta))^{(1-Y_{1t})(1-Y_{2t})}$$

With some abuse of notation, we also write $p(Y_t, \gamma)$. Define the identified sets for θ and θ_1 as:

$$\Theta_0 = \arg\sup_{\theta \in \Theta} E_0[\log p(Y_t, \theta)] \qquad \text{and} \qquad \Theta_{01} = \arg\sup_{\theta_1 \in \Theta_1} \sup_{\theta_2 \in \Theta_2(\theta_1)} E_0[\log p(Y_t, \theta)]$$

where the expectation E_0 is taken with respect to the true distribution of Y_t, P_0.

Once we fix V, Y_t depends only on the realisation of ε_t. Thus, under Assumption ID1 we can write the log-likelihood by $L_T(\theta) = \sum_{t=1}^T \log p(Y_t, \theta)$.[7] For a candidate value $\tilde{\theta}_1$ the profiled

[6] Otsu et al. (2016) develop tests for credibility of pooling markets in dynamic Markov games.

[7] If data come from N markets, replace T with NT everywhere (this assumes the same equilibrium is played in all markets).

likelihood ratio statistic is defined as:

$$LR_T(\tilde{\theta}_1) = 2\left[\sup_{\theta\in\Theta} L_T(\theta) - \sup_{\theta_2\in\Theta_2(\tilde{\theta}_1)} L_T(\tilde{\theta}_1, \theta_2)\right] + o_p(1) \tag{3}$$

where the $o_P(1)$ term accommodates approximation error coming from using a discrete grid for ε when calculating $\mathcal{V}$ sets (and optimization error). Similarly to Chen et al. (2018) we will calculate the $100\kappa\%$ confidence set for the identified set Θ_{01} as:

$$\widehat{\Theta}_{01}^{\kappa} = \{\theta_1 : LR_T(\theta_1) \le c_\kappa\}$$

where c_κ is the κ quantile of the asymptotic distribution of $\sup_{\theta_1\in\Theta_{01}} LR_T(\theta_1)$.

Obtaining asymptotic distribution of the likelihood ratio statistic in our case is difficult because the second supremum in (3) will often be attained at the boundary of the parameter space $\Theta_2(\tilde{\theta}_1)$ and the asymptotic distribution will depend on the shape of the local parameter space at this boundary (which will in turn depend on the value of θ_1). Given these difficulties we focus on obtaining an upper bound on the critical value.

Note that situation here is different than in the missing data example in Chen et al. (2018) as local parameter space at $\theta \in \Theta$ cannot be characterized as a simple translation of the null parameter space and, as a result, their Assumption 4.7 cannot be verified. Thus, we cannot use χ^2_1 quantile as an upper bound on our critical value.[8] Additionally, although their Procedure 2 is applicable in our model, it involves repeated computation of the level sets of the likelihood for each MCMC draw (see Appendix A.2. in their article) which is prohibitively costly in our setup as the likelihood is nonlinear and each evaluation requires recomputing the constraint set.

Let $\mathbb{I}_0$ denote the Fisher information matrix:

$$\mathbb{I}_0 = -E_0[\partial^2 \log p(y, \gamma_0)/\partial\gamma^2].$$

where $\gamma_0 = \gamma(\theta_0), \theta_0 \in \Theta_0$. We say that a sequence of sets Γ_T covers a set K if there is a sequence of closed balls of radius $k_T \to \infty$ centred at the origin, B_{k_T}, such that $\Gamma_T \cap B_{k_T} = K \cap B_{k_T}$ with

[8] Our Monte Carlo simulations (not reported here) also confirm that χ^2_1 critical value fails to provide required coverage.

probability approaching one (see p. 1987 in Chen et al. (2018)).

Assumption INF2. *We have:*

(a) Θ_1 *is a compact, nonempty subset of* $\mathbb{R}$ *and* $\Theta_2(\theta_1)$ *is a compact, convex, nonempty subset of* $\mathbb{R}^{d_2}$ *for every* $\theta_1 \in \Theta_1$.

(b) $\sup_{\theta_2 \in \Theta_2(\theta_1)} L_T(\theta_1, \theta_2)$ *is quasi-concave in* θ_1.

(c) $E_0[\partial \log p(y, \gamma)/\partial\gamma] = 0$ *has only one (interior) solution* $\gamma_0 = 0$.

(d) There exists a neighbourhood $\mathcal{N} \subset Int(\Gamma)$ *of* γ_0 *on which* $\log p(y, \gamma)$ *is twice continuously differentiable for each* y*, with first derivative in* $L^2(P_0)$*, and* $\sup_{\gamma \in \mathcal{N}} \|\partial^2 \log p(y, \gamma)/\partial\gamma^2\| \leq \bar{l}(y)$ *for* $\bar{l} \in L^2(P_0)$.

(e) There exists a neighbourhood $\mathcal{N}_\theta$ *of* Θ_0 *on which* $\gamma(\theta)$ *is twice continuously differentiable with second derivatives bounded uniformly over* $\mathcal{N}_\theta$.

(f) $\mathbb{I}_0$ *is non-singular and* $\mathbb{I}_0 = E_0\left[\frac{\partial \log p(y,\gamma)}{\partial\gamma}\frac{\partial \log p(y,\gamma)}{\partial\gamma}'\right]$.

(g) The local parameter space $\Gamma_{oT}(\theta_1) = \{\sqrt{T}\mathbb{I}_0^{1/2}\gamma(\theta_1, \theta_2) : (\theta_1, \theta_2) \in \Theta_{oT}\}$ *covers a closed convex cone* $\mathcal{K}(\theta_1) \subset \mathbb{R}^3$*, respectively, for every* $\theta_1 \in \Theta_1$*, where* Θ_{oT} *is a sequence of small neighbourhoods of* Θ_0.

Assumption (a) is satisfied in our setup because $\Theta_2(\theta_1) = [0, 1] \times \mathcal{V}(\theta_1)$ and $\mathcal{V}(\theta_1)$ is guaranteed to be convex (see Abreu & Sannikov (2014)).[9] Assumption (b) implies that $\sup_{\theta_1 \in \Theta_{01}} LR_T(\theta_1)$ is reached at the endpoints of the marginal identified set Θ_{01}. Denote these endpoints by $\underline{\theta}_1$ and $\overline{\theta}_1$. Assumptions (c)-(g) allow us to obtain quadratic approximation to the likelihood with respect to probabilities γ using Proposition 5.1 in Chen et al. (2018). They imply that the choice probabilities are identified from the population likelihood (but not θ).

Let $\Delta_{\theta_1} = \partial\gamma(\theta_0)/\partial\theta_2'$ for $\theta_0 = (\theta_1, \theta_2) \in \Theta_0$. In the simplest case, which is the leading case in our model, each value of the structural parameter θ_1 in the identified set Θ_{01} corresponds to a unique likelihood maximising value θ_2. This happens if the identified set is a manifold which is not parallel to any of the axes and implies that the θ_2 in the definition of Δ_{θ_1} is unique.

[9] The approximation to $\mathcal{V}$ dicussed in Section 4 remains convex as Minkowski sum preserves convexity.

Assumption INF3. *We have:*

(a) $\theta_2^*(\theta_1) = \arg\max_{\theta_2 \in \Theta_2(\theta_1)} L_T(\theta_1, \theta_2)$ *is a singleton for any* θ_1 *in the neighbourhood of* Θ_{01}.

(b) Δ_{θ_1} *has full row rank in the neighbourhood* $\mathcal{N}_\theta$.

A sufficient condition for part (a) is strict quasi-concavity of $L_T(\theta_1, \theta_2)$ in θ_2. Although this cannot always be easily verified analytically, problems with convergence of the numerical optimisation algorithms would usually signify violation of this condition. Part (b) is usually innocuous in our model as it merely requires that choice probabilities are not collinear in parameters. This assumption is similar to conditions in Lemma 1 in Kline & Tamer (2016).[10] If Assumption INF3 does not hold, one can still use the chi-square critical value described in Section 5.2 to obtain the desired (conservative) confidence set for θ_1.[11]

We say that the set Θ is approximated at θ_0 by a cone K^{θ_0} if:

$$\inf_{s \in K^{\theta_0}} \|(\theta - \theta_0) - s\| = o(\|\theta - \theta_0\|) \qquad \theta \in \Theta$$

$$\inf_{\theta \in \Theta} \|(\theta - \theta_0) - s\| = o(\|s\|) \qquad s \in K^{\theta_0}$$

Theorem 1. *Let* $\Theta_2(\underline{\theta}_1)$ *and* $\Theta_2(\overline{\theta}_1)$ *be approximated at* θ_2 *such that* $\theta_0 = (\underline{\theta}_1, \theta_2) \in \Theta_0$ *or* $\theta_0 = (\overline{\theta}_1, \theta_2) \in \Theta_0$ *by cones* $K_2(\underline{\theta}_1)$ *and* $K_2(\overline{\theta}_1)$, *respectively. If Assumptions INF1-INF3 hold, then:*

$$\sup_{\theta_1 \in \Theta_{01}} LR_T(\theta_1) = \max_{\theta_1 \in \{\underline{\theta}_1, \overline{\theta}_1\}} \left\{ \inf_{s \in K_2(\theta_1)} \|V_T - \Delta_{\theta_1} s\|^2 \right\} + o_p(1) \tag{4}$$

where $V_T = \mathbb{I}_0^{-1/2} \frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{\partial \log p(Y_t, \gamma_0)}{\partial \gamma}$ *asymptotically follows as a multivariate standard normal distribution.*

Theorem 1 shows that the asymptotic distribution of our statistic is equivalent to the distribution of the maximum of two chi-bar square distributions (cf. Proposition 3.4.1 in Silvapulle & Sen (2001)). Note that this bound is useful only if the probabilities $\gamma(\theta_0)$ are not flat in θ for

[10] Condition INF3(b) is akin to constraint qualification conditions in the stochastic programming literature. In fact, it is a localized version of the linear independence constraint qualification for equality constraints: $\gamma_0 - [p((1,1), \theta) \quad p((1,0), \theta) \quad p((0,1), \theta)]' = 0$, where the first component of θ is fixed at some θ_1 in the small neighbourhood of the identified set. See Kaido et al. (2020) for a discussion on how constraint qualifications relate to standard assumptions in the moment inequality literature.

[11] In Appendix E we discuss how to adjust our inference procedure in Section 5.1 if Assumption INF3 does not hold.

$\theta_0 = (\underline{\theta_1}, \theta_2)$ or $\theta_0 = (\overline{\theta_1}, \theta_2)$, which is guaranteed in our repeated game model (note that derivatives of γ involve density of the shocks, which is non-zero everywhere under standard assumptions).

5.1 Simulated critical value

Using Theorem 1 for inference on the identified set Θ_{01} presents both computational (calculation of constraint sets) and theoretical (parameter on boundary) challenges. For example, simply boostrapping our criterion function will not lead to a valid critical value. Thus, in order to provide a feasible inference procedure we make some simplifying assumptions and opt for a moderately conservative critical value.

5.1.1 Simplified $\mathcal{V}$ set

Firstly, note that using our likelihood criterion implies that we have to recalculate the $\mathcal{V}$ set, a convex polytope, for each candidate value of θ_1. This is computationally heavy, and barely feasible in realistic applications given the current state of computational resources. Thus, we simplify our procedure by enlarging the continuation payoff set $\mathcal{V}$ to a cube:

Assumption INF4. $\Theta_2(\theta_1)$ *is a cube in* $\mathbb{R}^{d_2}$*.*

Although this may result in a larger confidence set than implied by the general procedure in the previous section, it significantly simplifies computation of the critical value. Firstly, it allows us to track only the minimal and maximal values of the $\mathcal{V}$ set which considerably cuts the time needed to compute the profiled likelihood in (3). Secondly, it allows us to derive a simple upper bound on the asymptotic distribution described in Theorem 1 as now the local parameter space is an orthant in the least favourable case.

With this simplification the quantiles of the chi-bar square random variables described in Theorem 1 can be calculated using the formulas in Kudo (1963). However, the maximum in the formulation of the asymptotic approximation in (4) complicates obtaining the critical values as chi-bar square random variables under the maximum are correlated. Fortunately, with our extended $\mathcal{V}$ set we can derive a conservative critical value.

5.1.2 Bounding critical value from above

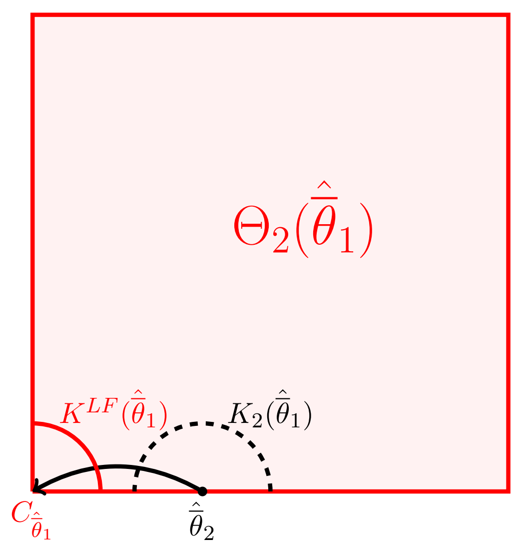
The first step in applying Theorem 1 lies in estimating Δ_{θ_1}, for which we need to estimate $\partial\gamma(\theta_0)/\partial\theta$ for $\theta_0 = (\underline{\theta_1}, \theta_2)$ and $\theta_0 = (\overline{\theta_1}, \theta_2)$. Thus, application of our results requires a preliminary estimate of the extreme points of the identified set Θ_0. We follow Chernozhukov et al. (2007) and suggest estimating these extreme points by:

$$
\begin{aligned}
(\hat{\underline{\theta}}_1, \hat{\underline{\theta}}_2):\quad & \hat{\underline{\theta}}_1 = \min\{\theta_1 : LR_T(\theta_1) \leq e_T\}; \hat{\underline{\theta}}_2 = \arg \sup_{\theta_2 \in \Theta_2(\hat{\underline{\theta}}_1)} L_T(\hat{\underline{\theta}}_1, \theta_2) \\
(\hat{\bar{\theta}}_1, \hat{\bar{\theta}}_2):\quad & \hat{\bar{\theta}}_1 = \max\{\theta_1 : LR_T(\theta_1) \leq e_T\}; \hat{\bar{\theta}}_2 = \arg \sup_{\theta_2 \in \Theta_2(\hat{\bar{\theta}}_1)} L_T(\hat{\bar{\theta}}_1, \theta_2)
\end{aligned}
$$

where $e_T = \log(\log(T))$ or $e_T = \log(T)$.[12]

The next step is to estimate the local cones $K_2(\underline{\theta_1}), K_2(\overline{\theta_1})$. This part is challenging as any sampling error in estimation of $\underline{\theta_1}$ may lead to incorrect estimate of $K_2(\underline{\theta_1})$. For example, even if $K_2(\underline{\theta_1})$ was an orthant in $\mathbb{R}^{d_2}$, $K_2(\hat{\underline{\theta}}_1)$ will likely be larger than an orthant, e.g. a half-space, even when T is relatively large.

Figure 3: Worst-case local parameter space



Note: The arcs mark local parameter spaces corresponding to different $\theta_2 \in \Theta_2(\hat{\bar{\theta}}_1)$

Instead, we choose to approximate K_2 by the closest least favourable local parameter space,

[12] If one believes that the degeneracy condition in Chernozhukov et al. (2007) is satisfied, one can use $e_T = 0$ or $e_T = o(1)$.

namely the orthant corresponding to the corner of the cube closest to $\underline{\hat{\theta}}_1$ or $\hat{\bar{\theta}}_1$.[13] We illustrate this construction in Figure 3 in a simplified setup in which $\Theta_2(\theta_1)$ is a square in $\mathbb{R}^2$ and Assumption INF3 holds. In this example $\hat{\bar{\theta}}_2$ lies on the edge of the parameter space so the local parameter space, $K_2(\hat{\bar{\theta}}_1)$, is a half-space $\mathbb{R} \times \mathbb{R}_+$. Instead of using $K_2(\hat{\bar{\theta}}_1)$, we take the local parameter space corresponding to the corner closest to $\hat{\bar{\theta}}_2$, namely $K^{LF}(\hat{\bar{\theta}}_1) = \mathbb{R}^2_+$.

This construction produces a critical value that, in general, will be larger than the critical value from the asymptotic distribution in Theorem 1. The intuitive reason for why the bound is only moderately conservative is the following. Consider the game Σ^S. In order to match the observed probabilities of observing (1,1) using a very low value of α we have to set the value contrasts D^1_{11} and D^2_{11} high. In other words, in order to entice players to cooperate on 1 when stage game payoffs from doing so are low we need to promise them high continuation values. As a result the lower and upper end of the identified set for α likely correspond to D's being in the corners of the feasible set for these contrasts.

Define $\mathcal{C}_{\theta_1}$ as the set of all corners of the cube $\Theta_2(\theta_1)$. Formally, we define our simulated critical value, c_κ, as the estimate of the κ quantile of the distribution of:

$$Q^{LF} \equiv \max_{\theta_1 \in \{\underline{\hat{\theta}}_1, \hat{\bar{\theta}}_1\}} \left\{ \inf_{s \in K^{LF}(C_{\theta_1})} \|V_T - \Delta_{\theta_1} s\|^2 \right\}$$

where $C_{\theta_1} \in \mathcal{C}_{\theta_1}$ is defined as:

$$C_{\theta_1} = \arg\min_{C \in \mathcal{C}_{\theta_1}} \|C - \theta_2^*(\theta_1)\|$$

for $\theta_2^*(\cdot)$ defined in Assumption INF3. $K^{LF}(C_{\theta_1})$ is an orthant in $\mathbb{R}^{d_2}$ approximating the parameter space $\Theta_2(\theta_1)$ at C_{θ_1}. Display 1 summarizes our procedure.[14]

[13]This resembles the geometric moment selection procedure in static entry games developed by Bontemps & Kumar (2020). However, their setup is very different than ours.

[14]An alternative bound on the critical value would come from setting $K^{LF}(\underline{\hat{\theta}}_1)$ to a cone with a smaller solid angle out of the conical hulls of $\Delta_{\underline{\hat{\theta}}_1}$ and $\Delta_{\hat{\bar{\theta}}_1}$ (least favourable cone) and setting its mirror image with respect to the origin as $K^{LF}(\hat{\bar{\theta}}_1)$ (least favourable alignment of the cones). This procedure produces similar but slightly more conservative critical values.

Display 1. *Simulation procedure:*

1. *Estimate* $\underline{\theta}_1$ *and* $\overline{\theta}_1$ *by:*

$$\hat{\underline{\theta}}_1 = \min\{\theta_1 : LR_T(\theta_1) \leq e_T\} \qquad \hat{\overline{\theta}}_1 = \max\{\theta_1 : LR_T(\theta_1) \leq e_T\}$$

   where $e_T = \log\log(T)$ *or* $e_T = \log(T)$*.*

2. *Estimate* $\Delta_{\underline{\theta}_1}$ *and* $\Delta_{\overline{\theta}_1}$ *by* $\Delta_{\hat{\underline{\theta}}_1} = \partial\gamma(\hat{\underline{\theta}}_1, \hat{\underline{\theta}}_2)/\partial\theta_2'$ *and* $\Delta_{\hat{\overline{\theta}}_1} = \partial\gamma(\hat{\overline{\theta}}_1, \hat{\overline{\theta}}_2)/\partial\theta_2'$*.*

3. *Find the corners,* $C_{\hat{\underline{\theta}}_1}$ *and* $C_{\hat{\overline{\theta}}_1}$*, of the parameter space closest to* $\hat{\underline{\theta}}_2$ *and* $\hat{\overline{\theta}}_2$*.*

4. *Simulate:*

$$\max\left\{\inf_{s\in K^{LF}(C_{\hat{\underline{\theta}}_1})} \|V - \Delta_{\hat{\underline{\theta}}_1} s\|^2, \inf_{s\in K^{LF}(C_{\hat{\overline{\theta}}_1})} \|V - \Delta_{\hat{\overline{\theta}}_1} s\|^2\right\}$$

   by drawing large number of standard normal vectors V*, and obtain the critical value* c_κ *by taking the* κ *quantile of the empirical distribution of this statistics across the simulation draws.*

5. *Estimate the* κ *confidence set for* θ_1 *by:*

$$\widehat{\Theta}_{01}^{\kappa} = \{\theta_1 : LR_T(\theta_1) \leq c_\kappa\}$$

Note that $\inf_{s\in K^{LF}(C_{\theta_1})} \|V_T - \Delta_{\theta_1} s\|^2$ is a convex program and can be solved repeatedly quite fast using packages like *CVX* in MATLAB. Additionally, the evaluations of $LR_T(\theta_1)$ in the first step of the procedure can be stored and reused to calculate the confidence set in the final step. Note that if we were to use Procedure 2 in Chen et al. (2018) we would have to calculate numerically the (marginal) level sets of $p(\cdot,(\theta_1,\cdot))$ with respect to the Kullback-Leibler distance for each MCMC draw from the posterior distribution of θ.[15] As it is not uncommon in the Bayesian estimation

[15] On top of that, the SMC generation of the posterior draws, described in Appendix A of Chen et al. (2018), involves repeated evaluation of the likelihood, which makes it even less attractive in a setup where such evaluations are costly, like the one in this article.

for the number of required MCMC draws to reach several thousand, application of that procedure is rather unattractive in our setup.[16] Comparing the results of Monte Carlo simulations in Chen et al. (2018) to those in Section 7 we conclude, however, that the simplified computation in our inference procedure comes at a cost of increased conservativeness, thus visualising the trade-off between computation and conservativeness that the researcher faces when choosing between these two methods.

We need to make the following technical assumptions in order to justify our procedure:

Assumption INF5. *C_{θ_1}, as a function of θ_1, is constant in θ_1 in the neighbourhoods of $\underline{\theta}_1$ and $\overline{\theta}_1$.*

This high-level assumption basically excludes a knife-edge situation in which $\underline{\theta}_2 = \theta_2^*(\underline{\theta}_1)$ and $\overline{\theta}_2 = \theta_2^*(\overline{\theta}_1)$ are (almost) equidistant from the corners of the parameter space and, as argued above, should be generally satisfied in our setup as the structure of our model suggests that $\underline{\theta}_2$ and $\overline{\theta}_2$ will be located at or close to the corners of the parameter space.

Theorem 2. *If assumptions of Theorem 1 and Assumptions INF3-INF5 hold, then for any $c \geq 0$:*

$$P\left(\sup_{\theta_1 \in \Theta_{01}} LR_T(\theta_1) \leq c\right) \geq P\left(\max_{\theta_1 \in \{\hat{\underline{\theta}}_1, \hat{\overline{\theta}}_1\}} \left\{\inf_{s \in K^{LF}(C_{\theta_1})} \|V_T - \Delta_{\theta_1} s\|^2\right\} \leq c\right) + o(1). \tag{5}$$

Theorem 2 confirms that our simulated critical value provides valid, but possibly conservative, inference for the identified set Θ_{01}. Our Monte Carlo simulations reported in Section 7 suggest that in fact this procedure is only mildly conservative, thus we recommend using it in applications.

Finally, observe that our bound on the critical value can also be applied in a static game, i.e. $\delta = 0$. Then, θ_2 corresponds to the static selection probability, π, and the optimisation problems involved in simulations become $\inf_{s \in \mathbb{R}_-} \|V_T - \Delta_{\hat{\theta}_1} s\|^2$ or $\inf_{s \in \mathbb{R}_+} \|V_T - \Delta_{\hat{\theta}_1} s\|^2$ as the worst case local parameter space for π is either $\mathbb{R}_-$ or $\mathbb{R}_+$. These problems can be solved analytically so there is no need for repeated numerical optimisation. Thus, our approximation may prove useful in the context of empirical static games when applying currently available bootstrap procedures (e.g. Kaido et al. (2019)) is computationally costly and assumptions listed above hold.[17]

[16] Note that the evaluation of $p(\cdot, (\theta_1, \cdot))$ for different draws of θ_1 can be run in parallel, which speeds up computation. Thus, the computational cost of this method may be manageable if one has access to a large number of computing nodes.

[17] Additionally, it applies in static games with box constraints on the payoff parameters α.

5.2 Simple conservative critical value

The procedure in the previous section, though feasible, still requires preliminary estimation of the identified set and a large number of convex optimisations. Therefore, in this section we suggest an alternative procedure that approximates the critical value from above. This bound is noticeably more conservative than the one implied by Theorem 2 but instead does not require simulation or additional computation in order to obtain the critical value.

Theorem 3. *If Assumption INF2 holds, then:*

$$\lim_{T\to\infty} P\left(\sup_{\theta_1\in\Theta_{01}} LR_T(\theta_1) \leq c_\kappa^\chi\right) \geq \kappa$$

where c_κ^χ is the κ quantile of the chi-square distribution with 3 degrees of freedom.

Theorem 3 is convenient because calculation of the critical value c_κ^χ is straightforward, though as shown in Section 7 this critical value gives much more conservative inference than our simulated critical value c_κ. In practice, one can compare confidence sets resulting from using both critical values for a chosen parsimonious specification of the model and apply the chi-square critical value for the remaining specifications only if the confidence sets do not significantly differ. Additionally, note that Theorem 3 does not require Assumption INF3, thus c_κ^χ can be applied in more general circumstances than c_κ.

The idea behind the conservative bound is similar to the conservative bound in Rosen (2008) as we basically assume that the constraints on the three probabilities $(\gamma_{11}(\theta), \gamma_{10}(\theta), \gamma_{01}(\theta))$, implied by the constraints on θ, are binding, though we have a likelihood model rather than a moment inequality model.

6 Model with covariates

The previous discussion did not explicitly include covariates. In practice, however, one may be interested in estimating the effect of observed characteristics of the agents and markets on the payoffs. For example, Ciliberto & Tamer (2009) estimate and simulate the effect of the Wright amendment on entry and exit into the Dallas airline markets. All the inference procedures described above can accommodate presence of covariates. However, doing so entails some conceptual and computational issues.

Firstly, this poses some conceptual dilemma as now the econometrician observes part of the stochastic component of the payoff vectors, $X = (X_1, X_2)$, hence she could potentially use history of the covariates to restrict the continuation payoff set $\mathcal{V}$, i.e. rule out some equilibria in the game. As this would prohibitively complicate the computation of the likelihood and it is not clear if this would allow us, in fact, to obtain much sharper bounds (note that the unobserved ε's have unbounded support), we assume that the econometrician does not use this knowledge, i.e. we treat X's just as ε's in the computation of the $\mathcal{V}$ set. Also we assume that the covariates are either i.i.d. or time-invariant (if we observe multiple markets/games in each period).

Secondly, the presence of covariates complicates computation – now $\mathcal{V}$ depends both on (X_1, X_2) and (α, β), where β is the vector of covariate coefficients. In order to deal with this complication we: 1) discretize (X_1, X_2) and 2) approximate the boundaries of $\mathcal{V}$, i.e. v_{min}, v_{max}, by a polynomial spline in (α, β). If we take as a point of departure the analysis of a partially-identified static game 1) can hardly be seen as a restriction as it is present also in the empirical analysis of that simple game (see Ciliberto & Tamer (2009)). Also our experience shows that the boundaries v_{min}, v_{max} tend to be smooth functions of (α, β) so the polynomial approximation works very well in practice.

Thirdly, now the "static" selection probability π may depend on X, which increases the dimensionality of the optimisation problems in our inference procedure. In order to address that, one may approximate $\pi(X_1, X_2)$ by an auxiliary parametric model, for example probit or logit (like in Bajari et al. (2010)).[18]

7 Monte Carlo simulations

7.1 Cournot entry game

In order to check the finite sample performance of our methods we performed a small Monte Carlo study using Cournot entry game example described above with independent shocks following logistic or Normal distribution $N(0, 4)$. As mentioned above, to simplify computation, we do not

[18] Here an attractive alternative may be using the moment inequality approach described in Appendix C as it does not involve the selection probability π.

calculate the full continuation payoff set $\mathcal{V}(\alpha)$ but only calculate the extreme values of this set: $v_{min} = \min\{v : v \in \mathcal{V}(\alpha)\}$ and $v_{max} = \max\{v : v \in \mathcal{V}(\alpha)\}$, and use $[v_{min}, v_{max}]^4$ in place of the continuation payoff set.

In simulations we set the true value $\alpha = 1$. The marginal identified set for α in the logistic model is $\Theta_{01} = [0.92, 1.1]$ for a discount factor $\delta = 0.55$ and $\Theta_{01} = [0.76, 1.89]$ for a discount factor $\delta = 0.75$.[19] For the normal model, these sets are $\Theta_{01} = [0.95, 1.09]$ and $\Theta_{01} = [0.8, 2.03]$. Results are given in Tables 1-2. We analyse performance of both the critical value implied by Theorem 2 (Simulated crit. val.) and Theorem 3 (χ^2_3 crit. val.).

The simulations confirm implications of our theoretical results – both critical values deliver coverage probabilities that are close or above nominal values. MC simulations confirm also that χ^2_3 critical value is a valid but, as expected, conservative upper bound on the critical value from the asymptotic distribution of our likelihood ratio statistic.

Our test has good power against values outside the identified set for $\delta = 0.55$ but for $\delta = 0.75$ it lacks power above the upper end of the identified set (especially with logistic shocks), which suggests that marginal likelihood is very flat at this end and resulting confidence sets will be quite wide.

7.2 Cournot entry game with covariates

We extend the Cournot entry game example to include observed covariates:

		P 2: 1	P 2: 0
P 1	1	$\alpha + X_1\beta - \varepsilon_1, \alpha + X_2\beta - \varepsilon_2$	$\frac{9}{4}\alpha + X_1\beta - \varepsilon_1, 0$
	0	$0, \frac{9}{4}\alpha + X_2\beta - \varepsilon_2$	$0, 0$

We draw both X_1 and X_2 from the uniform distribution over $\{-1, -0.5, 0, 0.5, 1\}$. As now the $\mathcal{V}$ set depends on α, β and (X_1, X_2), in order to simplify computation we generate a sample of v_{min} and v_{max} for given (X_1, X_2) and different values of α and β and approximate the relationship between (v_{min}, v_{max}) and α, β by a 6-th order bivariate polynomial. For simplicity, π is assumed to be the

[19] The maximization of the log-likelihood takes longer with larger δ, which is in line with the intuition that the likelihood should be flatter as the discount factor increases and the set of equilibrium values $\mathcal{V}$ grows (cf. Folk theorem). Thus, we do not perform full scale simulations for $\delta = 0.95$. Appendix D contains some partial results, which are similar to the ones here.

Table 1: MC simulations: coverage probabilities, $\delta = 0.55$

	Logistic shocks						Normal shocks						
	Simulated crit. val.			χ^2_3 crit. val.				Simulated crit. val.			χ^2_3 crit. val.		
	90%	95%	99%	90%	95%	99%		90%	95%	99%	90%	95%	99%
$T = 100$													
$\Theta_{01} = [0.92, 1.1]$	0.966	0.985	0.995	0.989	0.995	0.999	$\Theta_{01} = [0.95, 1.09]$	0.952	0.984	0.997	0.99	1	1
$\alpha = 0.5$	0.037	0.074	0.231	0.108	0.193	0.387	$\alpha = 0.5$	0.027	0.056	0.192	0.08	0.15	0.39
$\alpha = 1.5$	0.724	0.839	0.937	0.881	0.925	0.978	$\alpha = 1.6$	0.243	0.321	0.552	0.41	0.51	0.7
$T = 250$													
$\Theta_{01} = [0.92, 1.1]$	0.916	0.978	0.997	0.989	0.996	1	$\Theta_{01} = [0.95, 1.09]$	0.931	0.965	0.995	0.97	0.99	1
$\alpha = 0.5$	0.001	0.001	0.001	0.001	0.001	0.008	$\alpha = 0.5$	0	0	0.001	0	0	0.01
$\alpha = 1.5$	0.512	0.653	0.808	0.702	0.776	0.893	$\alpha = 1.6$	0.025	0.041	0.103	0.06	0.09	0.21
$T = 500$													
$\Theta_{01} = [0.92, 1.1]$	0.923	0.973	0.986	0.98	0.983	0.997	$\Theta_{01} = [0.95, 1.09]$	0.942	0.969	0.992	0.98	0.99	1
$\alpha = 0.5$	0	0	0	0	0	0	$\alpha = 0.5$	0	0	0	0	0	0
$\alpha = 1.5$	0.321	0.425	0.615	0.457	0.601	0.767	$\alpha = 1.6$	0	0.001	0.003	0	0	0.01
$T = 1000$													
$\Theta_{01} = [0.92, 1.1]$	0.929	0.975	0.995	0.978	0.994	0.998	$\Theta_{01} = [0.95, 1.09]$	0.939	0.966	0.993	0.98	0.99	1
$\alpha = 0.5$	0	0	0	0	0	0	$\alpha = 0.5$	0	0	0	0	0	0
$\alpha = 1.5$	0.116	0.156	0.313	0.217	0.299	0.473	$\alpha = 1.6$	0	0	0	0	0	0
$T = 2000$													
$\Theta_{01} = [0.92, 1.1]$	0.933	0.966	0.992	0.981	0.991	0.999	$\Theta_{01} = [0.95, 1.09]$	0.941	0.967	0.994	0.98	0.99	1
$\alpha = 0.5$	0	0	0	0	0	0	$\alpha = 0.5$	0	0	0	0	0	0
$\alpha = 1.5$	0.007	0.014	0.054	0.025	0.045	0.118	$\alpha = 1.6$	0	0	0	0	0	0

Note: 2000 Monte Carlo replications

Table 2: MC simulations: coverage probabilities, $\delta = 0.75$

	Logistic shocks							Normal shocks					
	Simulated crit. val.			χ^2_3 crit. val.				Simulated crit. val.			χ^2_3 crit. val.		
	90%	95%	99%	90%	95%	99%		90%	95%	99%	90%	95%	99%
$T = 100$													
$\Theta_{01} = [0.76, 1.89]$	0.981	0.99	0.996	0.994	0.995	0.999	$\Theta_{01} = [0.8, 2.03]$	0.974	0.985	0.994	0.99	0.99	1
$\alpha = 0.5$	0.226	0.356	0.592	0.413	0.485	0.811	$\alpha = 0.3$	0	0.001	0.004	0	0	0.02
$\alpha = 2.25$	0.937	0.938	0.985	0.938	0.983	0.996	$\alpha = 2.25$	0.749	0.903	0.952	0.91	0.91	0.97
$T = 250$													
$\Theta_{01} = [0.76, 1.89]$	0.895	0.991	0.999	0.999	0.999	1	$\Theta_{01} = [0.8, 2.03]$	0.966	0.975	0.997	0.98	0.98	1
$\alpha = 0.5$	0.017	0.034	0.129	0.053	0.106	0.282	$\alpha = 0.3$	0	0	0	0	0	0
$\alpha = 2.25$	0.801	0.881	0.949	0.914	0.914	0.968	$\alpha = 2.25$	0.585	0.745	0.868	0.75	0.87	0.93
$T = 500$													
$\Theta_{01} = [0.76, 1.89]$	0.922	0.973	0.984	0.98	0.983	0.999	$\Theta_{01} = [0.8, 2.03]$	0.941	0.972	0.988	0.98	0.99	1
$\alpha = 0.5$	0	0	0.001	0.001	0.001	0.013	$\alpha = 0.3$	0	0	0	0	0	0
$\alpha = 2.25$	0.691	0.767	0.899	0.818	0.906	0.955	$\alpha = 2.25$	0.38	0.49	0.662	0.52	0.65	0.76
$T = 1000$													
$\Theta_{01} = [0.76, 1.89]$	0.931	0.972	0.992	0.977	0.992	0.997	$\Theta_{01} = [0.8, 2.03]$	0.962	0.979	0.997	0.99	1	1
$\alpha = 0.5$	0	0	0	0	0	0	$\alpha = 0.3$	0	0	0	0	0	0
$\alpha = 2.25$	0.421	0.573	0.748	0.689	0.689	0.858	$\alpha = 2.25$	0.118	0.173	0.323	0.2	0.28	0.47
$T = 2000$													
$\Theta_{01} = [0.76, 1.89]$	0.936	0.965	0.994	0.982	0.993	0.999	$\Theta_{01} = [0.8, 2.03]$	0.966	0.983	0.996	0.99	1	1
$\alpha = 0.5$	0	0	0	0	0	0	$\alpha = 0.3$	0	0	0	0	0	0
$\alpha = 2.25$	0.202	0.275	0.452	0.369	0.447	0.63	$\alpha = 2.25$	0.007	0.017	0.047	0.02	0.05	0.1

Note: 2000 Monte Carlo replications

same for different values of (X_1, X_2). Otherwise the MC designs are the same as in the simple model above.

We report only the results for $\delta = 0.75$ (see Table 3). The marginal identified set for α in the logistic model is $\Theta_{01} = [0.33, 3.25]$ and $\Theta_{01} = [0.38, 2.78]$ in the normal model. The results confirm previous conclusions, namely that our simulation method is not excessively conservative, in fact it produces coverage probabilities that are close to the nominal values, which is encouraging. Using the chi-squared with three degrees of freedom is far more conservative, just as in the model without covariates. Overall, the results in Table 3 confirm that introducing covariates into the model does not essentially change the properties of our inference procedures.

8 Empirical illustration: Wright Amendment

We employ our framework to re-evaluate the Wright Amendment experiment in Ciliberto & Tamer (2009) (CT henceforth). The Wright Amendment, passed into law in 1979, restricted airline service from Dallas Love airport in order to stimulate growth of Dallas/Forth Worth, permitting flights only to/from Texas, Louisiana, Arkansas, Oklahoma, New Mexico, Alabama, Kansas and Mississippi. The Amendment was partially repealed starting from 2006 and fully withdrawn in 2014. Using a static game model CT compare the predicted changes in market service out of Dallas Love with and without the amendment and find very large positive effects of repealing the amendment on the number of served markets. As the repeal is predicted to especially benefit Southwest Airlines the authors conclude that the prolonged binding of the Amendment was meant to protect American Airlines monopolies in markets out of Dallas/Fort Worth.

They interpret these effects as short run effects because their model does not involve dynamic dimension. However, dynamic strategic responses are important in this context. Firstly, a repeated game model without Markov assumption allows collusive behaviour, which helps to explain $(1, 1)$, $(1, 0)$ and $(0, 1)$ outcomes in the game. Secondly, changes in the market conditions will affect firms decisions to be present in the market and the dynamic model takes that into account. In fact, as Table 4 shows, there is a lot of entry and exit in the US airline market across time. For example, Delta dropped out of 42% of the markets they served in 1990, even more strikingly, in 2010 they came back to 9.5% of their 1990 markets after not serving them in 2000. Thus, the network of

Table 3: MC simulations: coverage probabilities, $\delta = 0.75$, model with covariates

	Logistic shocks							Normal shocks					
	Simulated crit. val.			χ^2_3 crit. val.				Simulated crit. val.			χ^2_3 crit. val.		
	90%	95%	99%	90%	95%	99%		90%	95%	99%	90%	95%	99%
$T = 100$													
$\Theta_{01} = [0.33, 3.25]$	0.958	0.986	0.998	0.991	0.995	1	$\Theta_{01} = [0.38, 2.78]$	0.937	0.971	0.998	0.98	0.99	1
$\alpha = 0.05$	0.281	0.41	0.675	0.463	0.594	0.813	$\alpha = 0.05$	0.301	0.426	0.693	0.51	0.64	0.83
$\alpha = 3.75$	0.873	0.927	0.976	0.944	0.969	0.992	$\alpha = 3.25$	0.844	0.909	0.969	0.94	0.96	0.99
$T = 250$													
$\Theta_{01} = [0.33, 3.25]$	0.915	0.968	0.998	0.98	0.993	1	$\Theta_{01} = [0.38, 2.78]$	0.91	0.96	0.993	0.98	0.99	1
$\alpha = 0.05$	0.033	0.063	0.178	0.09	0.156	0.334	$\alpha = 0.05$	0.041	0.08	0.214	0.12	0.19	0.38
$\alpha = 3.75$	0.696	0.788	0.91	0.84	0.895	0.965	$\alpha = 3.25$	0.646	0.757	0.892	0.81	0.88	0.96
$T = 500$													
$\Theta_{01} = [0.33, 3.25]$	0.909	0.95	0.993	0.978	0.99	0.999	$\Theta_{01} = [0.38, 2.78]$	0.907	0.95	0.99	0.97	0.99	1
$\alpha = 0.05$	0	0.002	0.007	0.003	0.006	0.025	$\alpha = 0.05$	0.001	0.002	0.011	0	0.01	0.04
$\alpha = 3.75$	0.454	0.566	0.76	0.644	0.74	0.878	$\alpha = 3.25$	0.399	0.52	0.722	0.59	0.69	0.84
$T = 1000$													
$\Theta_{01} = [0.33, 3.25]$	0.921	0.958	0.994	0.98	0.992	1	$\Theta_{01} = [0.38, 2.78]$	0.907	0.951	0.984	0.97	0.98	0.99
$\alpha = 0.05$	0	0	0	0	0	0	$\alpha = 0.05$	0	0	0	0	0	0
$\alpha = 3.75$	0.179	0.265	0.462	0.336	0.436	0.651	$\alpha = 3.25$	0.124	0.195	0.404	0.27	0.37	0.58
$T = 2000$													
$\Theta_{01} = [0.33, 3.25]$	0.922	0.966	0.994	0.983	0.992	1	$\Theta_{01} = [0.38, 2.78]$	0.92	0.963	0.992	0.98	0.99	1
$\alpha = 0.05$	0	0	0	0	0	0	$\alpha = 0.05$	0	0	0	0	0	0
$\alpha = 3.75$	0.018	0.037	0.108	0.054	0.092	0.211	$\alpha = 3.25$	0.01	0.021	0.069	0.03	0.06	0.15

Note: 2000 Monte Carlo replications

served markets is far from stable and carriers often react to changing market conditions by exit and entry.

Table 4: Share of markets by presence of major carriers over time (in %)

Presence in Q2 1990 - Q2 2000 - Q2 2010	American	Delta	United	US Airways	Southwest
in - in - in	21.7	24.8	26.1	13.9	59.9
in - out - in	5.6	9.5	3.3	5.3	15.8
in - out - out	60.3	42.2	50.4	56.5	17.3
in - in - out	12.5	23.5	20.3	24.3	7.1

Note: T100 Market data, flights with fewer than 20 passengers in 1990 dropped. Markets are defined as directional routes between origin and destination airports (irrespective of the number of stops on the way). "In" means that a carrier served at least one flight on the route.

In order to address these concerns we redo the experiment using our repeated game model focusing on the interaction between American (AA) and Southwest (WN) in the markets out of Dallas. We use quarterly DB1B and T100 data from 1993 to 2017. We estimate our model using the following covariates: *market size*, the size of the market calculated as the geometric mean of the population of the endpoint cities, *Wright*, indicating that the market was restricted by the Wright amendment, *airport presence*, which gives the average fraction of markets served from the endpoints by American or Southwest and *cost* which is equal to the difference between the "origin-closest hub-destination" distance and the nonstop "origin-destination" distance divided by the latter distance (see CT for detailed description and justification).

Table 5: Summary statistics

	Mean	Min	Max
market size	3,675,950	1,750,056	10,968,557
Wright	0.356	0	1
AA airport presence	0.763	0.211	0.983
WN airport presence	0.367	0	0.713
AA cost	0.009	0	0.083
WN cost	0.238	0	2.811

Table 5 contains the summary statistics. There are 244 markets in our sample.[20] The average market covered a population of around 3-4 million people. As we focus on Dallas, the Wright amendment affects a large fraction of markets – 35.6% of year-quarter-market observations in our

[20]Note that we assume that the same repeated game is played in each of these markets, thus we focus on, what we believe to be, a homogenous group of Dallas markets.

sample were affected by this law. The average AA airport presence is significantly higher and AA cost significantly lower than in CT sample as they focus on a larger set of markets and only use data from the second quarter of 2001. These numbers show that AA had a strong presence at Dallas airports and that it mainly operated direct flights to all the destinations outside Dallas, which is in line with the fact that Dallas has been traditionally a major hub for AA.

8.1 Estimation results

We estimate the confidence sets using our profiled likelihood criterion with simulated critical value described in Section 5.1.1. We draw 1000 random standard normal vectors in order to obtain this critical value. The stage game in our model has the form:

		WN: 1	WN: 0
AA	1	$X_1\beta - \varepsilon_1, X_2\beta - \varepsilon_2$	$\alpha + X_1\beta - \varepsilon_1, 0$
	0	$0, \alpha + X_2\beta - \varepsilon_2$	$0, 0$

Thus, we restrict both the competitive effect (α) and the effect of covariates (β) to be the same for AA and WN.[21] We also assume that the selection probability π does not depend on (X_1, X_2), i.e. if two equilibria are present in the normal form game in markets characterised by (X_1, X_2), firms play one of them with the same probability as in markets characterised by $(\tilde{X}_1, \tilde{X}_2) \neq (X_1, X_2)$. We discretize all continuous variables as binary (below/above mean) and include a constant in both X_1 and X_2. We consider both strong ($\delta = 0.75$) and weak ($\delta = 0.95$) discounting of future payoffs by the carriers and compare the results to the static case ($\delta = 0.00$), i.e. when static game equilibrium is played repeatedly. Table 6 contains the confidence sets estimated using our procedure.[22]

We observe that the data provides useful information about the underlying model parameters. In particular, we can identify the sign of all coefficients besides the competitive effect and the cost variable. Our estimates are also in the same ballpark as the estimates obtained using the static game approach in CT. Our model, however, does not contain a lot of information about the competitive

[21] The motivation for using homogenous coefficients is computational. Note that allowing α and β to vary between AA and WN would require evaluating the equilibrium continuation correspondence $\mathcal{V}$ for each candidate set of values $(\alpha_{AA}, \alpha_{WN}, \beta_{AA}, \beta_{WN})$, which significantly increases computational burden compared to the homogenous case.

[22] As in CT, routes from Dallas Love and Dallas/Fort Worth are separate markets so separate games are played in these markets and AA cannot use DFW routes to fight WN Love routes.

Table 6: Estimated confidence sets from repeated game with random states, Dallas market

	$\delta = 0.75$		$\delta = 0.95$		$\delta = 0.00$	
Parameter	90% CS	95% CS	90% CS	95% CS	90% CS	95% CS
comp. effect (α)	[-37.26, 33.92]	[-37.26, 33.92]	[-60.34 , 33.64]	[-60.34, 33.64]	[-0.2, -0.01]	[-0.22, 0.00]
market size	[0.34, 0.42]	[0.32, 0.43]	[0.33, 0.43]	[0.32, 0.44]	[0.32, 0.35]	[0.31, 0.36]
Wright	[-0.65, -0.54]	[-0.66, -0.53]	[-0.65, -0.54]	[-0.67, -0.53]	[-1.38, -1.33]	[-1.39, -1.32]
airport presence	[4.57, 4.67]	[4.56, 4.68]	[4.54, 4.7]	[4.53, 4.71]	[4.4, 4.46]	[4.39, 4.46]
cost	[-0.01, 0.09]	[-0.03, 0.11]	[-0.02, 0.1]	[-0.03, 0.11]	[-0.05, 0.12]	[-0.07, 0.14]
constant	[3.55, 3.96]	[3.51, 4]	[-3.42, 3.37]	[-3.59, 3.37]	[-2.95, -2.89]	[-2.95, -2.89]

Note: Critical values for confidence sets calculated using simulated critical value described in Display 1 with 1000 random Normal draws. The sets were built using a grid search with step 0.01. The dataset contains 244 markets out of/to Dallas.

effect as the estimated bounds are very wide and include both positive and negative values.[23] Interestingly, the bounds on other coefficients are very tight, suggesting that these parameters are close to being point-identified. As expected, the bounds obtained for the discount factor $\delta = 0.95$ are, in general, wider than those for $\delta = 0.75$ which corresponds to the fact that the set of possible equilibrium continuation values $\mathcal{V}$ expands as the discount factor grows. Though, the difference in the width of the confidence set is only pronounced for the competitive effect α, with bounds for the near-to-point-identified parameters being very close for both discount factors.

Comparing the results to the static case, we can see that allowing multiple equilibria in the dynamic dimension comes at the cost of much wider confidence intervals for the competitive effect (and the constant). With $\delta = 0.00$ we can identify the sign of α, similarly to CT. Interestingly, imposing lack of dynamics leads to much more negative estimates of the effect of the Wright amendment than in the general model, which is also in line with CT findings about the strong (or even too strong) effect of the amendment on restricting competition on the Dallas airline market in the static model. What seems to happen is that, in order to explain the difference in frequencies of outcome $(1, 0)$ in the "non-Wright" markets versus "Wright" markets (65% of the former are "AA only" with 4% of the latter), the static model estimates a large effect of the Wright amendment. Since the dynamic model can partly explain high frequencies of $(1, 0)$ in non-restricted markets by collusive equilibria, it produces more reasonable estimates of the effect of the Wright amendment.

The fact that the bounds for the Wright coefficient with $\delta = 0.00$ are not contained in the bounds for $\delta = 0.75$ or $\delta = 0.95$ suggests that imposing a static equilibrium in each period may be

[23] Note that the 90% and 95% confidence sets for α are the same. This is because we use a 0.01 step in the grid search for building the confidence sets. The log-likelihood is steep around the CS endpoints so the difference in the endpoints of 90% and 95% sets is less than 0.01.

refuted by the data. This shows that combining repeated games theory with data puts meaningful restrictions on the equilibria that might have been played in our dynamic game.

8.2 Policy experiment

Finally, we use our estimated confidence sets to simulate the effect of repealing the Wright amendment from 2006. We look at the change in predicted probabilities in our model: p(1,1), p(1,0), p(0,1), p(0,0), with the actual policy and the scenario in which the Wright amendment is in operation throughout the sample period 1993-2017. In order to obtain the confidence sets for the policy effects: 1) we take all possible combinations of extreme points of the confidence sets , 2) we randomly draw 10000 parameter values from the estimated confidence sets, and for each point selected in these ways we calculate the change in probabilities.[24] Then we take the union of the sets of estimated policy effects from 1) and 2). As in CT, we focus on markets out of Dallas Love airport, which were restricted by the amendment. We use the 95% confidence sets from Table 6.

Table 7: Estimated effect of repealing the Wright amendment, 95% level

		Change in probability (Wright - non Wright)	
		$\delta = 0.75$	$\delta = 0.00$
Both AA and WN	p(1,1)	[-0.177, 0]	[-0.042, -0.029]
Only AA	p(1,0)	[-0.138, 0.163]	[-0.311, -0.192]
Only WN	p(0,1)	[-0.048, 0.116]	[-0.004, -0.002]
Not served	p(0,0)	[0, 0.103]	[0.234, 0.344]

Note: The change in probabilities is calculated as the predicted probability assuming that the Wright amendment is in place throughout the sample period 1993-2017 minus the predicted probability with actual trajectory of *Wright*. The other variables are kept at their sample values.

Table 7 shows that keeping the Wright amendment in place would most likely decrease the number of markets served by both AA and WN, even by up to 17.7%, and increase the number of non-served markets by up to 10.3%. However, the results do not provide a decisive prediction for what would happen with markets served only by AA or only by WN, with the possibility that keeping Wright amendment in place would increase the number of these markets. These results suggest that keeping the Wright amendment in force might have restricted competition in significant number of markets, preventing entry of competitive airlines.

[24] For each configuration of the parameters we take $\theta_2 = (\pi, V)$ that maximizes the log-likelihood (see Assumption INF3).

Comparing our estimated policy effects with $\delta = 0.75$ and the static case of $\delta = 0.00$, we see that the static model predicts much stronger effect of keeping the amendment in place on the number of unserved markets than the dynamic model. However, it also predicts that this increase in the number of unserved markets would come mainly at the cost of markets served only by AA, with unambiguously negative effect of Wright amendment on the number of markets served only by this carrier. Although we use a different dataset and a different model, our results with $\delta = 0.00$ are similar to the ones in Section 6 of CT. CT predict even a 63.84% drop in the number of non-served markets after the repeal of the amendment, with possibly up to 47.44% of these markets served by American and/or Southwest. We see that our dynamic model predicts much more modest, and thus more plausible, effect of the Wright amendment.

The static game model with $\delta = 0.00$ will generally have trouble generating collusive enter-enter equilibria even without Wright restrictions. This explains a very small effect of lifting the Wright amendment on the probability of both AA and WN entering observed in Table 7. In contrast, the dynamic model facilitates appearance of collusive equilibria after lifting the amendment which shows itself in a strong and unambiguous increase in the number of markets served by both carriers as a result of this policy change.

Overall, our results in this section illustrate the usefulness of our approach to analysing repeated strategic interactions between firms and show that combining the repeated games model with the data may lead to interesting findings and improved analysis of strategic interactions in the US airline market.

9 Discussion

Although we focus on repeated games with random states, our identification and inference approach can be extended to general stochastic games with state dependence. Recent advancements in the analysis of these games (Abreu et al. (2016), Abreu et al. (2020)) provide readily available procedures for computing equilibrium payoff sets, which can be naturally embedded into our econometric approach. Additionally, the assumption of binary action sets made in our exposition can be relaxed as the computation algorithm in Abreu & Sannikov (2014) allows non-binary actions and our inference procedure in Display 1 applies in this case, though it is important to note that the computational burden increases with the dimension of the action space as dimension of the continuation vector V grows exponentially with $|\mathcal{A}|$.

As we allow a wide range of possible equilibria, the confidence sets may be quite large in empirical applications. One may restrict the number of possible equilibria in these games by refining the equilibrium concept, e.g. focusing on strategy-proof equilibria, which should result in smaller continuation payoff sets $\mathcal{V}$ and narrower bounds on the parameters of interest.

We focus on entry decisions in the dynamic context in our paper, thus we leave out the pricing decisions. Goolsbee & Syverson (2008) show that airlines may adjust prices in advance when faced with the threat of entry, thus adding pricing to the model would be an important extension. See Ciliberto et al. (2021) for a recent effort to model entry and pricing decisions at the same time using a static game.

Appendix

A Mathematical Proofs

A.1 Proof of Theorem 1

First note that under quasi-concavity:

$$\inf_{\theta_1 \in \Theta_{01}} \sup_{\theta_2 \in \Theta_2(\theta_1)} L_T(\theta_1, \theta_2) = \min_{\theta_1 \in \{\underline{\theta}_1, \overline{\theta}_1\}} \sup_{\theta_2 \in \Theta_2(\theta_1)} L_T(\theta_1, \theta_2) \tag{6}$$

Further, Assumptions INF2(c)-(g) allow us to apply Proposition 5.1 and Lemma F.1 in Chen et al. (2018) in order to obtain quadratic expansion of the likelihood in γ:

$$\begin{aligned}\sup_{\theta \in \Theta_{oT}} 2L_T(\theta) &= 2\sum_{t=1}^{T} \log p(Y_t, \gamma_0) + \|V_T\|^2 - \inf_{\theta \in \Theta_{oT}} \|\sqrt{T}\mathbb{I}_0^{1/2}\gamma(\theta) - V_T\|^2 + o_p(1) = \\ &= 2\sum_{t=1}^{T} \log p(Y_t, \gamma_0) + \|V_T\|^2 + o_p(1)\end{aligned}$$

where $V_T = \mathbb{I}_0^{-1/2} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial \log p(Y_t, \gamma_0)}{\partial \gamma}$. Now for the restricted part:

$$\begin{aligned}\sup_{\theta \in \Theta_{oT}(\theta_1)} 2L_T(\theta) &= 2\sum_{t=1}^{T} \log p(Y_t, \gamma_0) + \|V_T\|^2 - \inf_{\theta = (\theta_1, \theta_2) \in \Theta_{oT}} \|\sqrt{T}\mathbb{I}_0^{1/2}\gamma(\theta) - V_T\|^2 + o_p(1) = \\ &= 2\sum_{t=1}^{T} \log p(Y_t, \gamma_0) + \|V_T\|^2 - \inf_{\kappa \in \Gamma_{oT}(\theta_1)} \|\kappa - V_T\|^2 + o_p(1) = \\ &= 2\sum_{t=1}^{T} \log p(Y_t, \gamma_0) + \|V_T\|^2 - \inf_{\kappa \in \mathcal{K}(\theta_1)} \|\kappa - V_T\|^2 + o_p(1)\end{aligned}$$

This implies:

$$LR_T(\theta_1) = 2\left[\sup_{\theta \in \Theta_{oT}} L_T(\theta) - \sup_{\theta \in \Theta_{oT}(\theta_1)} L_T(\theta)\right] = \inf_{\kappa \in \mathcal{K}(\theta_1)} \|\kappa - V_T\|^2 + o_p(1)$$

Now use Assumption INF3 and proceed similarly to Shapiro (1985). Note that the cone $\mathcal{K}(\theta_1)$ can be approximated by:

$$\left\{\frac{\partial\gamma(\theta_0)}{\partial\theta_2'}s : s \in K_2(\theta_1), \theta_0 = (\theta_1, \theta_2) \in \Theta_0\right\} = \{\Delta_{\theta_1}s : s \in K_2(\theta_1)\}$$

for $\theta_1 \in \{\underline{\theta_1}, \overline{\theta_1}\}$, which implies that:

$$LR_T(\theta_1) = \inf_{s\in K_2(\theta_1)} \|V_T - \Delta_{\theta_1}s\|^2 + o_p(1)$$

and together with (6) concludes the proof.

A.2 Proof of Theorem 2

First we will demonstrate that $\hat{\underline{\theta}}_1 \to^p \underline{\theta}_1$ and $\hat{\overline{\theta}}_1 \to^p \overline{\theta}_1$ as $T \to \infty$. For this purpose we apply Theorem 3.1 in Chernozhukov et al. (2007). Their condition C.1 is satisfied as follows: part (a) follows from our Assumption INF2(a), lower-semicontinuity of $E[\log p(Y_t, \theta)]$ in part (b) holds in the neighbourhood $\mathcal{N}_\theta$ of Θ_0 by our Assumptions INF2(d) and INF2(e), part (c) follows from our continuity assumptions on $p(Y_t, \theta)$ and discreteness of Y_t, uniform convergence in part (d) can be shown to hold over $\mathcal{N}_\theta$ by applying Jennrich's ULLN with the help of our Assumptions INF2(a),(d),(e) and noting that $0 \le p(Y_t, \theta) \le 1$.

Next recall that B_{k_T} denotes a ball centred at zero with radius $k_T \to \infty$ and define $KB_{k_T}(\theta_1) = \{\kappa : \kappa = \Delta_{\theta_1}s, s \in B_{k_T}\}$. If $\Delta_{\theta_1} = 0$, then trivially $\inf_{s\in K^{LF}(\theta_1)} \|V_T - \Delta_{\theta_1}s\|^2 = \inf_{s\in K^{LF}(\theta_1)\cap B_{k_T}} \|V_T - \Delta_{\theta_1}s\|^2$ so we focus on the case when $\Delta_{\theta_1} \neq 0$. Note that in this case $KB_{k_T}(\theta_1)$ is an ellipsoid. For θ_1 in the neighbourhoods of $\underline{\theta}_1$ and $\overline{\theta}_1$ we have:

$$P\left(\inf_{s\in K^{LF}(\theta_1)} \|V_T - \Delta_{\theta_1}s\|^2 - \inf_{s\in K^{LF}(\theta_1)\cap B_{k_T}} \|V_T - \Delta_{\theta_1}s\|^2 \neq 0\right) \le P(V_T \notin KB_{k_T}(\theta_1))$$

but as the elliptic radi grow with k_T we have that $P(V_T \notin KB_{k_T}(\theta_1)) \to 0$ as $T \to \infty$. Thus we can write:

$$\inf_{s\in K^{LF}(\theta_1)} \|V_T - \Delta_{\theta_1}s\|^2 = \inf_{s\in K^{LF}(\theta_1)\cap B_{k_T}} \|V_T - \Delta_{\theta_1}s\|^2 + o_p(1) \tag{7}$$

Assumption INF5 implies that $K^{LF}(\theta_1) \cap B_{k_T}$ is a continuous (compact-valued) correspondence around $\underline{\theta}_1$ and $\overline{\theta}_1$. Additionally Δ_{θ_1} is continuous in θ_1 in this neighbourhood by Assumption INF2(e). Thus, we can apply Berge's maximum theorem to conclude that $\inf_{s \in K^{LF}(\theta_1) \cap B_{k_T}} \|V_T - \Delta_{\theta_1} s\|^2$ is a continuous function of θ_1. Now (7) and continuous mapping theorem imply:

$$\begin{aligned} \inf_{s \in K^{LF}(\hat{\underline{\theta}}_1) \cap B_{k_T}} \|V_T - \Delta_{\hat{\underline{\theta}}_1} s\|^2 &= \inf_{s \in K^{LF}(\underline{\theta}_1)} \|V_T - \Delta_{\underline{\theta}_1} s\|^2 + o_p(1) \\ \inf_{s \in K^{LF}(\hat{\overline{\theta}}_1) \cap B_{k_T}} \|V_T - \Delta_{\hat{\overline{\theta}}_1} s\|^2 &= \inf_{s \in K^{LF}(\overline{\theta}_1)} \|V_T - \Delta_{\overline{\theta}_1} s\|^2 + o_p(1) \end{aligned} \tag{8}$$

as $T \to \infty$.

Next note that $K^{LF}(\theta_1) \subseteq K_2(\theta_1)$. This is trivially satisfied when $K_2(\theta_1) = \mathbb{R}^{d_2}$ or $K_2(\theta_1)$ is an orthant itself. Consider the remaining case when $K_2(\theta_1) = \mathbb{R}_+^{d_+} \times \mathbb{R}_-^{d_-} \times \mathbb{R}^{d_2 - d_+ - d_-}$ where $0 \leq d_+ + d_- \leq d_2 - 1$. Now we must have $K^{LF}(\theta_1) = \mathbb{R}_+^{d_+} \times \mathbb{R}_-^{d_-} \times \mathbb{R}^{\tilde{d}_+} \times \mathbb{R}^{\tilde{d}_-}$ with $\tilde{d}_+ + \tilde{d}_- = d_2 - d_+ - d_-$. To see that, without loss of generality, suppose that $K^{LF}(\theta_1) = \mathbb{R}_- \times \mathbb{R}_+^{d_+ - 1} \times \mathbb{R}_-^{d_-} \times \mathbb{R}^{\tilde{d}_+} \times \mathbb{R}^{\tilde{d}_-}$ and let $\tilde{C}$ be the corner associated with this parameter space and $\theta_2 = \theta_2^*(\theta_1)$ be the profiled-likelihood-minimising value. Now note that the first coordinate of $\tilde{C}$, $\tilde{C}_1$, has to be different than the first coordinate of θ_2, $\theta_{2,1}$, but these are the same for the closest corner, i.e. $C_{\theta_1,1} = \theta_{2,1}$. We have:

$$\|\tilde{C} - \theta_2\|^2 = |\tilde{C}_1 - \theta_{2,1}|^2 + \|\tilde{C}_{-1} - \theta_{2,-1}\|^2 > \inf_{C \in \mathcal{C}} \|C_{-1} - \theta_{2,-1}\|^2 = \|C_{\theta_1} - \theta_2\|^2$$

which implies that $\tilde{C}$ cannot be the closest corner to θ_2.

Finally, we have:

$$\max\left\{ \inf_{s \in K^{LF}(\underline{\theta}_1)} \|V_T - \Delta_{\underline{\theta}_1} s\|^2, \inf_{s \in K^{LF}(\overline{\theta}_1)} \|V_T - \Delta_{\overline{\theta}_1} s\|^2 \right\} \geq \max\left\{ \inf_{s \in K_2(\underline{\theta}_1)} \|V_T - \Delta_{\underline{\theta}_1} s\|^2, \inf_{s \in K_2(\overline{\theta}_1)} \|V_T - \Delta_{\overline{\theta}_1} s\|^2 \right\}$$

which together with (8) concludes the proof.

A.3 Proof of Theorem 3

For a closed convex cone K let K^o denote its polar cone (see e.g. Section 14 in Rockafellar (1970)). From the proof of Theorem 1 and Moreau's decomposition theorem (note that $\mathcal{K}(\theta_1)$ is a closed convex cone by Assumption INF2(g)):

$$LR_T(\theta_1) = \|V_T\|^2 - \inf_{\kappa \in \mathcal{K}^o(\theta_1)} \|\kappa - V_T\|^2 + o_p(1)$$

which implies $LR_T(\theta_1) \leq \|V_T\|^2 + o_p(1)$ and

$$\sup_{\theta_1 \in \Theta_{01}} LR_T(\theta_1) \leq \max_{\theta_1 \in \{\underline{\theta}_1, \overline{\theta}_1\}} \{\|V_T\|^2, \|V_T\|^2\} + o_p(1) = \|V_T\|^2 + o_p(1)$$

and the result follows from V_T being asymptotically $N(0, I)$.

B Alternative definitions of market presence

In Table 4 an airline is present in the market if it operates at least one flight from the market origin to market destination. Here we consider other definitions of market presence. First, we use DB1B ticketing data, which contains 10% sample of airline tickets from reporting carriers, and redefine market presence as selling at least 5 tickets for the specified route (see Table 8). Next, we use T100 Segment data and redefine market as a segment of the trip, for example a flight from ORD to MIA through DCA contains two segments ORD-DCA and DCA-MIA (using previous definition this would only be a single market ORD-MIA).

Table 8: Share of markets by presence of major (ticketing) carriers over time (in %)

Presence in Q2 1993 - Q2 2002 - Q2 2012	American	Delta	United	US Airways	Southwest
in - in - in	43.4	61.1	74.8	42.8	86.8
in - out - in	9	5.9	9.2	12.3	4.8
in - out - out	37	23.1	12.8	27	6
in - in - out	10.6	9.9	3.2	17.9	2.4

Note: DB1B Market data, flights with less than 5 tickets in 1993 dropped. Markets are defined as directional routes between origin and destination airports (irrespective of the number of stops on the way). "In" means that a carrier served at least one flight on the route.

The numbers in Table 8 significantly differ from those in Table 4 as ubiquitous codeshare and interlining agreements drive a wedge between the definitions of operating and ticketing carrier. The differences are smaller between Table 9 and Table 4. Despite these differences the main message remains the same – there is substantial amount of entry and exit across time in the US airline market.

Table 9: Share of markets by presence of major carriers over time (in %)

Presence in Q2 1990 - Q2 2000 - Q2 2010	American	Delta	United	US Airways	Southwest
in - in - in	34	35.8	38.6	16.5	74.6
in - out - in	5	6.8	2.7	3.5	3.1
in - out - out	46.4	33.4	35.8	53.8	14
in - in - out	14.6	24.1	23	26.3	8.3

Note: T100 Segment data, flights with less than 20 passengers in 1990 dropped. Markets are defined as segments of directional routes between origin and destination airports. "In" means that a carrier served at least one flight on the route.

C Moment inequality characterisation

For a non-empty $A \subsetneq \mathcal{A}$ let $\mathcal{L}(A;\alpha,V)$ denote the probability of observing some $a \in A$ in equilibrium in the normal form game under the assumption that in the regions of multiple equilibria an equilibrium in A is always selected. For simplicity let Assumption INF1 hold. Following Galichon & Henry (2011) the marginal identified set for α can be characterised using moment inequalities by:

$$\Theta_{01}^{S} = \{\alpha : E_0(\mathbb{1}\{a \in A_j\}) \leq \mathcal{L}(A_j;\alpha,V), A_j \subsetneq \mathcal{A}, V \in \mathcal{V}_S(\alpha)\} \tag{9}$$

where $j = 1, 2, \ldots, J$. Now for inference one can implement either the profiled procedure in Kaido et al. (2019) (KMS) or Bugni et al. (2017) (BCS). We discuss how the computational burden of these procedures compares to our approach as computation is the main obstacle for a practical inference in our model.[25] For this discussion we employ similar notation as in Section 5 in the paper, namely $\theta_1 \equiv \alpha$, $\theta_2 \equiv V$, $\theta = (\theta_1, \theta_2)$.

C.1 KMS inference

Let $G_{n,j}^{b}$ be a standardised estimator of $E_0(\mathbb{1}\{a \in A_j\})$ evaluated on a bootstrap sample scaled by $\sqrt{n}$ and let $\hat{D}_{n,j}(\theta)$ denote the gradient of $\mathcal{L}(A_j;\alpha,V)$ w.r.t. α and V normalised by the sample

[25] To be clear, both KMS and BCS apply to a much wider range of problems than our approach. Thus we only argue that for our specialised model the simulated critical value described in the paper may be preferable to KMS and BCS from a practical perspective.

standard deviation of moment j, $\hat{\sigma}_{n,j}$. Further, let $\hat{\xi}_{n,j}(\theta)$ denote $(\iota_n \hat{\sigma}_{n,j})^{-1}\sqrt{n}$ times the sample estimator of moment j, where $\iota_n \to \infty$. The KMS critical value is obtained by bootstrapping:

$$\Lambda_n^b(\theta, \rho, c) = \{\lambda \in \sqrt{n}(\Theta - \theta) \cap \rho B^d : G_{n,j}^b + \hat{D}_{n,j}(\theta)\lambda + \psi_j(\hat{\xi}_{n,j}(\theta)) \leq c, j = 1, 2, \ldots, J\}$$

where ρB^d imposes a technical "box" constraint on the local parameter space and ψ_j is a Generalised Moment Selection function of Andrews & Soares (2010), and can be calculated as:

$$\hat{c}(\theta) = \inf\{c \in \mathbb{R}_+ : P^*(\Lambda_n^b(\theta, \rho, c) \cap \{\lambda_1 = 0\} \neq \emptyset) \geq \kappa\}$$

where P^* denotes the law induced by bootstrap sampling and λ_1 is the first element of λ. Finally, the marginal confidence set is built by finding lowest and highest value of θ_1 for which the sample moment inequalities are satisfied with slackness $\hat{c}(\theta_1, \theta_2)$.

Let us now compare our inference method to KMS. Note that the computationally difficult step in our model is the re-evaluation of the continuation value set $\Theta_2(\theta_1)$ for different values of θ_1, which will be embedded in evaluating $\sqrt{n}(\Theta - \theta)$ within $\Lambda_n^b(\theta, \rho, c)$ in the KMS procedure. Our procedure in Display 1 controls the number of evaluations of $\Theta_2(\theta_1)$ by controlling the size of the grid for candidate values of θ_1 in the pre-estimation of the identified set in Step 1 and re-using evaluations of the likelihood ratio from Step 1 in building the confidence set in the final Step 5.

Similarly, the first step in the KMS procedure in which candidate values of θ are drawn can be adjusted to include only θ's on the grid of values for θ_1 to limit number of evaluations of $\Theta_2(\theta_1)$. However, as currently implemented, the KMS procedure proceeds with a smooth iterative algorithm to generate further "good" candidate values of θ ("A-M steps") and, thus, requires recalculation of $\Theta_2(\theta_1)$ for these newly generated values. As the number of iterations required for this algorithm to converge may differ from application to application, it may be difficult to control the number of evaluations of $\Theta_2(\theta_1)$ in practice without significant changes to this algorithm. Therefore, it seems that using the moment inequality characterisation in (9) and KMS is unlikely to dominate our method in terms of computational convenience.

C.2 BCS inference

The main BCS critical value is obtained by taking a minimum over two profiled bootstrap statistics, $T_n^{DR}(\theta_1)$ and $T_n^{PR}(\theta_1)$, in order to improve power. As our inference procedure is conservative it seems fair to compare it to $T_n^{DR}(\theta_1)$ and $T_n^{PR}(\theta_1)$ separately rather than to the more computationally intensive minimum statistic.

Using the notation from the previous section BCS resampling statistics can be written as:

$$
\begin{aligned}
T_n^{DR}(\theta_1) &= \inf_{\theta_2 \in \widehat{\Theta}_2(\theta_1)} \sum_{j=1}^{J} \left[G_{n,j}^b + \psi_j(\hat{\xi}_{n,j}(\theta)) \right]_- \\
T_n^{PR}(\theta_1) &= \inf_{\theta_2 \in \Theta_2(\theta_1)} \sum_{j=1}^{J} \left[G_{n,j}^b + \hat{\xi}_{n,j}(\theta) \right]_-
\end{aligned}
$$

where $\hat{\Theta}_2(\theta_1)$ is the set of minimizers of the KMS test statistic (see their paper for details). Note that similarly to our approach we can control the number of evaluations of $\Theta_2(\theta_1)$ by imposing a grid on θ_1. However, note that resampling both $T_n^{DR}(\theta_1)$ and $T_n^{PR}(\theta_1)$ requires repeatedly solving a non-linear non-convex constrained optimisation problem. Also, as argued in the main text, solution to this problem will often be reached on the boundary of the set $\widehat{\Theta}_2(\theta_1)$ and $\Theta_2(\theta_1)$.[26] This is much more computationally expensive than repeatedly solving a convex optimisation problem in our simulation procedure in Display 1.

D Additional Monte Carlo simulations

We perform limited number of simulations for $\delta = 0.95$ due to slow convergence of optimization algorithms for this case, which leads to extensive computing times. The results are given in Table 10.

[26] If we impose a condition equivalent of Assumption INF3, namely we have $\widehat{\Theta}_2(\theta_1) = \{\theta_2^*(\theta_1)\}$ the DR statistic seems computationally attractive compared to our approach. Still it requires evaluation of $\theta_2^*(\theta_1)$ for all the candidate values θ_1 whereas we only need to calculate it for the endpoints of the pre-estimated marginal identified set, $\hat{\underline{\theta}}_1$ and $\hat{\bar{\theta}}_1$.

Table 10: MC simulations: coverage probabilities, $\delta = 0.95$

	Normal shocks					
	Simulated crit. val.			χ^2_3 crit. val.		
	90%	95%	99%	90%	95%	99%
$\Theta_{01} = [0.49, 5.06]$	0.979	0.981	0.981	0.981	0.981	0.985
$\alpha = 0.1$	0	0	0	0	0	0
$\alpha = 5.6$	0.268	0.270	0.273	0.270	0.270	0.276
$\Theta_{01} = [0.49, 5.06]$	0.92	0.971	0.98	0.971	0.973	0.996
$\alpha = 0.1$	0	0	0	0	0	0
$\alpha = 5.6$	0.131	0.132	0.135	0.133	0.135	0.135
$\Theta_{01} = [0.49, 5.06]$	0.935	0.961	0.982	0.982	0.982	0.994
$\alpha = 0.1$	0	0	0	0	0	0
$\alpha = 5.6$	0.084	0.085	0.088	0.085	0.087	0.088
$\Theta_{01} = [0.49, 5.06]$	0.942	0.97	0.989	0.971	0.985	0.994
$\alpha = 0.1$	0	0	0	0	0	0
$\alpha = 5.6$	0.076	0.077	0.079	0.077	0.077	0.079

Note: 500 Monte Carlo replications

E Inference without Assumption INF3

In this section we discuss how we can adjust our simulated critical value if $\theta^*(\theta_1)$ is not unique and $\partial\gamma(\theta_0)/\partial\theta_2'$ contains zero rows for some θ_0. This will happen, for example, if the true probabilities, γ_0, are flat on a set with non-empty interior.

Firstly, the main complication here comes from the fact that now the local parameter space for γ at γ_0 cannot be approximated simply by taking $\left\{\frac{\partial\gamma(\theta_0)}{\partial\theta_2'}s : s \in K_2(\theta_1), \theta_0 = (\theta_1, \theta_2) \in \Theta_0\right\}$ (cf. proof of Theorem 1) as this set maps to zero when Θ_0 is a compact set with non-empty interior. However, given that γ is smooth around γ_0 (see Assumption INF2(e)) a simple expansion $\Theta_{0\eta} = \{\theta \in \Theta : \inf_{\theta_0 \in \Theta_0} \|\theta - \theta_0\| \leq \eta\}$ for $\eta > 0$ would allow us to bound the asymptotic distribution of our profiled criterion as:

$$\sup_{\theta_1 \in \Theta_{01}} LR_T(\theta_1) \leq \max_{\theta_1 \in \{\underline{\theta}_1, \overline{\theta}_1\}} \left\{ \inf_{\kappa \in \bigcup_{(\theta_1,\theta_2)\in\Theta_{0\eta}} \left\{\frac{\partial\gamma(\theta_1,\theta_2)}{\partial\theta_2'}s : s \in K_2^{\theta_2}(\theta_1)\right\}} \|V_T - \kappa\|^2 \right\} + o_p(1)$$

$$= \max_{\theta_1 \in \{\underline{\theta}_1, \overline{\theta}_1\}} \left\{ \inf_{(\theta_1,\theta_2)\in\Theta_{0\eta}} \inf_{s \in K_2^{\theta_2}(\theta_1)} \left\| V_T - \frac{\partial\gamma(\theta_1,\theta_2)}{\partial\theta_2'}s \right\|^2 \right\} + o_p(1)$$

where $K_2^{\theta_2}(\theta_1)$ is a cone approximating the local parameter space at $\theta_2 \in \theta_2^*(\theta_1)$.

Now in order to approximate the statistic on the right-hand side above, we can replace $\underline{\theta}_1$ and $\overline{\theta}_1$ with $\hat{\underline{\theta}}_1$ and $\hat{\overline{\theta}}_1$ as before. Next note that the identified set estimator in Chernozhukov et al. (2007) approximates the identified set from "outside", thus in practice we can replace $\Theta_{0\eta}$ with their estimator $\hat{\Theta}_0$ (note that we only have to estimate a "slice" out of $\hat{\Theta}_0$ for $\underline{\hat{\theta}}_1$ and $\hat{\overline{\theta}}_1$). Finally, note that now the (set of) closest corner(s) to $\theta_2 \in \theta_2^*(\theta_1)$ depends on the value of θ_2, $C_{(\theta_1,\theta_2)}$, and Assumption INF5 is too strong in this setup. Thus, letting $K^{LF}(C)$ denote the orthant corresponding to the corner $C \in C_{(\theta_1,\theta_2)}$ we can replace $\inf_{s \in K_2^{\theta_2}(\theta_1)} \left\| V_T - \frac{\partial \gamma(\theta_1,\theta_2)}{\partial \theta_2'} s \right\|^2$ above with $\max_{C \in C_{(\theta_1,\theta_2)}} \inf_{s \in K^{LF}(C)} \left\| V_T - \frac{\partial \gamma(\theta_1,\theta_2)}{\partial \theta_2'} s \right\|^2$ in order to simulate a conservative critical value.

Acknowledgments: We would like to thank Mike Abito, Wojciech Olszewski, Subir Bose and Andre Stenzel for useful discussions and suggestions and to Arpita Ghosh and Shuai Zhao for providing competent research assistance. We are grateful to participants in the LSE Joint Econometrics and Statistics Workshop, the (Ce)2 Workshop, the UEA Economics seminar and various conferences for useful comments. This research used the ALICE High Performance Computing Facility at the University of Leicester.

References

Abito, J. M. & Chen, C. (2021), 'How much can we identify from repeated games?', *Economics Bulletin* **41**(3), 1212–1222.

Abreu, D., Brooks, B. & Sannikov, Y. (2016), A "pencil-sharpening" algorithm for two player stochastic games with perfect monitoring. Working Paper.

Abreu, D., Brooks, B. & Sannikov, Y. (2020), 'Algorithms for stochastic games with perfect monitoring', *Econometrica* **88**(4), 1661–1695.

Abreu, D., Pearce, D. & Stacchetti, E. (1990), 'Toward a theory of discounted repeated games with imperfect monitoring', *Econometrica* **58**(5), pp. 1041–1063.

Abreu, D. & Sannikov, Y. (2014), 'An algorithm for two-player repeated games with perfect monitoring', *Theoretical Economics* **9**(2), 313–338.

Al Mohamad, D., Van Zwet, E. W., Cator, E. & Goeman, J. J. (2020), 'Adaptive critical value for constrained likelihood ratio testing', *Biometrika* **107**(3).

Amemiya, T. (1986), *Advanced econometrics*, Blackwell, Oxford.

Andrews, D. W. K. (2001), 'Testing when a parameter is on the boundary of the maintained hypothesis', *Econometrica* **69**(3), 683–734.

Andrews, D. W. K. & Soares, G. (2010), 'Inference for parameters defined by moment inequalities using generalized moment selection', *Econometrica* **78**(1), 119–157.

Bajari, P., Benkard, C. L. & Levin, J. (2007), 'Estimating dynamic models of imperfect competition', *Econometrica* **75**(5), 1331–1370.

Bajari, P., Hong, H. & Ryan, S. P. (2010), 'Identification and estimation of a discrete game of complete information', *Econometrica* **78**(5), 1529–1568.

Beresteanu, A., Molchanov, I. & Molinari, F. (2011), 'Sharp identification regions in models with convex moment predictions', *Econometrica* **79**(6), 1785–1821.

Berry, S. & Jia, P. (2010), 'Tracing the woes: An empirical analysis of the airline industry', *American Economic Journal: Microeconomics* **2**(3), 1–43.

Bontemps, C. & Kumar, R. (2020), 'A geometric approach to inference in set-identified entry games', *Journal of Econometrics* **218**(2).

Bugni, F. A., Canay, I. A. & Shi, X. (2017), 'Inference for subvectors and other functions of partially identified parameters in moment inequality models', *Quantitative Economics* **8**(1), 1–38.

Chen, X., Christensen, T. M. & Tamer, E. (2018), 'Monte Carlo confidence sets for identified sets', *Econometrica* **86**(6), 1965–2018.

Chernozhukov, V., Hong, H. & Tamer, E. (2007), 'Estimation and confidence regions for parameter sets in econometric models', *Econometrica* **75**(5), 1243–1284.

Ciliberto, F., Murry, C. & Tamer, E. (2021), 'Market structure and competition in airline markets', *Journal of Political Economy* **129**(11), 2995–3038.

Ciliberto, F. & Tamer, E. (2009), 'Market structure and multiple equilibria in airline markets', *Econometrica* **77**(6), 1791–1828.

Galichon, A. & Henry, M. (2011), 'Set identification in models with multiple equilibria', *The Review of Economic Studies* **78**(4), 1264–1298.

Gerardi, K. & Shapiro, A. (2009), 'Does competition reduce price dispersion? New evidence from the airline industry', *Journal of Political Economy* **117**(1), 1–37.

Giacomini, R. & Kitagawa, T. (2021), 'Robust bayesian inference for set-identified models', *Econometrica* **89**(4), 1519–1556.

Goolsbee, A. & Syverson, C. (2008), 'How do incumbents respond to the threat of entry? Evidence from the major airlines', *The Quarterly Journal of Economics* **123**(4), 1611–1633.

Kaido, H., Molinari, F. & Stoye, J. (2019), 'Confidence intervals for projections of partially identified parameters', *Econometrica* **87**, 1397–1432.

Kaido, H., Molinari, F. & Stoye, J. (2020), Constraint qualifications in partial identification. Working Paper.

Kline, B. & Tamer, E. (2016), 'Bayesian inference in a class of partially identified models', *Quantitative Economics* **7**(2), 329–366.

Kudo, A. (1963), 'A multivariate analogue of the one-sided test', *Biometrika* **50**(3/4), 403–418.

Lee, B. S. & Stewart, C. (2016), 'Identification of payoffs in repeated games', *Games and Economic Behavior* **99**, 82–88.

Magnac, T. & Thesmar, D. (2002), 'Identifying dynamic discrete decision processes', *Econometrica* **70**(2), pp. 801–816.

Mailath, G. & Samuelson, L. (2006), *Repeated Games and Reputations: Long-Run Relationships*, Oxford University Press.

Otsu, T., Pesendorfer, M. & Takahashi, Y. (2016), 'Pooling data across markets in dynamic Markov games', *Quantitative Economics* **7**(2), 523–559.

Rockafellar, R. T. (1970), *Convex analysis*, Princeton Mathematical Series, Princeton University Press.

Romano, J. P. & Shaikh, A. M. (2008), 'Inference for identifiable parameters in partially identified econometric models', *Journal of Statistical Planning and Inference* **138**(9), 2786–2807.

Rosen, A. M. (2008), 'Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities', *Journal of Econometrics* **146**(1), 107–117.

Salz, T. & Vespa, E. (2020), 'Estimating dynamic games of oligopolistic competition: an experimental investigation', *The RAND Journal of Economics* **51**(2), 447–469.

Shapiro, A. (1985), 'Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints', *Biometrika* **72**(1), 133–144.

Silvapulle, M. J. & Sen, P. K. (2001), *Constrained Statistical Inference*, John Wiley & Sons, Inc.

Tamer, E. (2003), 'Incomplete simultaneous discrete response model with multiple equilibria', *The Review of Economic Studies* **70**(1), 147–165.