COUNTERFACTUALS WITH LATENT INFORMATION

By

Dirk Bergemann, Benjamin Brooks, and Stephen Morris

January 2019
Revised October 2021

COWLES FOUNDATION DISCUSSION PAPER NO. 2162R4

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Counterfactuals with Latent Information[*]

Dirk Bergemann[†] Benjamin Brooks[‡] Stephen Morris[§]

October 1, 2021

We describe a methodology for making counterfactual predictions in settings where the information held by strategic agents and the distribution of payoff-relevant states of the world are unknown. The analyst observes behavior assumed to be rationalized by a Bayesian model, in which agents maximize expected utility, given partial and differential information about the state. A counterfactual prediction is desired about behavior in another strategic setting, under the hypothesis that the distribution of the state and agents' information about the state are held fixed. When the data and the desired counterfactual prediction pertain to environments with finitely many states, players, and actions, the counterfactual prediction is described by finitely many linear inequalities, even though the latent parameter, the information structure, is infinite dimensional.

Keywords: Counterfactuals, Bayes correlated equilibrium, information structure, linear program.

JEL Classification: C72, D44, D82, D83.

[*] We acknowledge financial support through NSF Grants SES 1459899, SES 1824137 and SES 2049744. We are grateful for valuable discussions with Stéphane Bonhomme, Hanming Fang, Phil Haile, Tibor Heumann, Peter Hull, Lorenzo Magnolfi, Eduardo Souza-Rodrigues, Philipp Strack, Vasilis Syrgkanis, Elie Tamer and Ed Vytlacil; and excellent research assistance from Aroon Narayanan and Roi Orzach. The paper has also benefited tremendously from the comments of the editor Jeff Ely and four anonymous referees.

[†] Department of Economics, Yale University, dirk.bergemann@yale.edu

[‡] Department of Economics, University of Chicago, babrooks@uchicago.edu

[§] Department of Economics, Massachusetts Institute of Technology, semorris@mit.edu

1 Introduction

A central problem in the empirical analysis of strategic settings is the presence of incomplete information. For example, in an auction, it matters not only what bidders believe about their value from the object if they win it, but also what they believe about others' beliefs about the value of the object to them and to others. In making predictions about outcomes in a counterfactual strategic environment, it is standard and natural to hold fixed the distribution of payoff-relevant states and players' beliefs about the state. These are collectively described by an *information structure.* Because informational structures are complex, most applied work relies on strong simplifying assumptions, such as independent private values in an auction. But the strength of these assumptions undermines the credibility of the resulting counterfactual predictions.

In this paper, we provide a novel approach to counterfactual analysis with latent and fixed information. We describe a completely non-parametric approach to identifying counterfactuals by treating the information structure as a nuisance parameter—that is, a parameter which is not of intrinsic interest but needs to be accounted for in making counterfactual predictions. We thus avoid the complexity that would be involved in identifying the information structure itself.[1] By instead focusing on a given counterfactual of interest and finite-dimensional restrictions on information implicit in observed behavior, we are able to derive a concise description of the set of counterfactual predictions.

Let us describe our approach in a simple case. Suppose that an individual takes an action a from a finite set of possible actions. The action results in a payoff $u(a, \theta)$, where θ is an uncertain state of the world that has finitely many possible values. Suppose that we observe the empirical distribution of (a, θ), which we refer to as the *outcome* and denote by $\phi(a, \theta)$. We do not observe the individual's information about θ, but we maintain the standard

[1] In an earlier version of this paper (Bergemann, Brooks, and Morris, 2021), we discuss in greater detail in Section 7.1 the conceptual difficulties associated with identifying the information structure. At the very least, we would need to identify a distribution over the players' infinite sequences of higher-order beliefs.

assumption that the action maximizes expected utility, given whatever beliefs about θ are held at the time the decision was made.

Now suppose we wish to predict how the same representative individual will behave in a new decision problem, where an action $\widehat{a}$ leads to a payoff $\widehat{u}\left(\widehat{a},\theta\right)$. Importantly, while the decision problem changes, the joint distribution of θ and the individual's information about θ are assumed to remain the same. Under the assumption that θ is observed, the distribution of θ can be computed directly from ϕ, but the distribution of beliefs conditional on θ is a latent parameter. Our question is: which counterfactual outcomes $\widehat{\phi}\left(\widehat{a},\theta\right)$ could be induced by optimal behavior for some information structure which also rationalizes the observed data ϕ?

Our main result implies the following simple and tractable characterization of these predictions: Imagine that rather than performing an abstract thought experiment, the individual actually did choose $\widehat{a}$ at the same time as a was chosen, and we simply did not observe it. The payoff was simply the sum of the payoffs across the two decision problems, so that there was no interaction between the two choices except through the common information. We refer to this as the *linked decision problem*, in which the action is an ordered pair $(a,\widehat{a})$. Since both actions were taken based on the same information about the same state, there will be correlation between θ, a, and $\widehat{a}$, and we can write $\overline{\phi}\left(a,\widehat{a},\theta\right)$ for the joint distribution of these objects. For $\overline{\phi}$ to be consistent with our data, the marginal on (a,θ) must be ϕ. The counterfactual prediction is simply the marginal of $\overline{\phi}$ on $(\widehat{a},\theta)$.

Thus, the problem of computing counterfactual predictions can be reduced to computing those outcomes $\overline{\phi}$ for the linked problem which are consistent with Bayesian rationality. But Bergemann and Morris (2013, 2016) have shown that these are precisely the Bayes correlated equilibria (BCE) of the linked problem, which are a convex set of $\overline{\phi}$ that satisfy a finite collection of "obedience constraints" that encode Bayesian optimality. When we add in the constraint that the marginal on (a,θ) is the observed ϕ, we obtain a convex set of joint distributions $\overline{\phi}$ on $(a,\widehat{a},\theta)$ which are consistent with rationality in the linked decision problem and are consistent with the data. The marginals of these distributions on $(\widehat{a}, \theta)$ is precisely the set of counterfactual outcomes that are consistent with the data. When there are finitely many actions and states and the data is given by finitely many linear constraints, counterfactual predictions are defined by a finitely many linear inequalities, which one can use, for example, to compute the range of counterfactual welfare via linear programming.

This outline presumed that there was one player and that the entire outcome was observed. Our subsequent analysis shows that this logic goes through in general many-player finite-action games, where the data reveals linear constraints on ϕ. In particular, our methodology applies to situations in which we only observe the players' actions, or some statistic thereof, such as the winning bid in an auction. In this case, the state is unobserved, and in our characterization of counterfactual predictions, the prior is implicitly restricted by the data. The argument that counterfactual predictions are characterized by linear inequalities is completely general, and follows the same steps as our informal discussion above.

We present two examples that illustrate our characterization and show how features of the information structure are encoded in the data. The first is a discrete version of Roy's (1951) model of self selection: a worker decides whether or not to join the workforce (the action) in the presence of uncertainty about potential earnings (the state). We first consider both the case where the entire outcome is observed, so that we see potential earnings for all workers, and also the case where the data is censored and potential earnings are only observed conditional on employment. The example shows how the observed correlation between employment and wages restricts the worker's information, which in turn determines counterfactual welfare when the potential earnings rise or fall. The second example is a two-firm entry game in the spirit of Ciliberto and Tamer (2009). The novelty in this strategic setting is that the observed outcome restricts both what the each firm knows about the state, and also what they know about the other firm's information.

Following the examples, we discuss some straightforward extensions that are likely to be important for applications. Our main result is presented in a stark theoretical benchmark where the analyst knows places no a priori restrictions on players' information. There are many natural settings where an analyst would know more than in our baseline model but less than in standard parameterized models of the information structure. In Section 5, we show how assumptions about the information structure can be incorporated in order to tighten the set of counterfactual predictions. For example, the analyst might be confident that players know those features of the state that affect their own preferences, which we refer to as the case of private values; or the analyst may observe "payoff shifters" that affect a player's preferences but are orthogonal to other elements of the model.

As noted above, our approach builds on the work of Bergemann and Morris (2013, 2016) who showed that BCE are precisely the outcomes that can arise under some information structure and equilibrium, for a fixed game. The contribution of this paper is to introduce linked games to characterize counterfactual predictions when information is the same in the observed and counterfactual games. A different approach to counterfactuals building on BCE results is to hold the prior fixed but impose no restriction on how players' information about the state might vary between the observations and counterfactual. This "unrestricted-information" approach was discussed informally in Bergemann and Morris (2013) and Bergemann et al. (2017) and implemented by Magnolfi and Roncoroni (2021) and Syrgkanis et al. (2021).[2] We discuss this and other ways of relaxing the fixed-information assumption in Section 5 and discuss these and other econometric papers in Section 6.

Before proving our main result, we derive a foundational result on joint predictions across games when there is a common information structure, in terms of BCE of a linked game. Our result on counterfactuals follows when we consider two games, one of which generates some observable data and the other of which is the counterfactual. The joint predictions result has potential applications beyond counterfactuals. In Section 7, we argue that it can be used to derive "informationally robust" rankings of games, which are illustrated in the Online Appendix. Heumann (2019) analyzed a version of this problem in the context

[2] In recent versions of their work, Magnolfi and Roncoroni (2021) study fixed-information counterfactuals as well as unrestricted-information counterfactuals

of symmetric games with normal uncertainty and linear best responses. The methods and results are complementary. Our approach is more general, which expands the range of potential applications but may make it harder to obtain analytic results.

The rest of this paper proceeds as follows. Section 2 establishes the basic notation. Section 3 presents our main results on counterfactual predictions. Section 4 presents our examples. Section 5 presents extensions. Section 6 discusses related literature. Section 7 concludes the paper. An Online Appendix contains additional results and examples.

2 Preliminaries

There is a state of the world $\theta \in \Theta$, where Θ is finite. There is also a set of players, indexed by $i = 1, \ldots, N$. The players and state space will be held fixed throughout our analysis.

The players interact through a *base game* $\mathcal{G} = (A_i, u_i)_{i=1}^N$, where A_i is the set of actions of player i, $A = \times_{i=1}^N A_i$ is the set of action profiles, and $u_i : A \times \Theta \to \mathbb{R}$ represents player i's expected utility preferences over (a, θ).

The distribution of the state and players' beliefs are described by a common-prior information structure $\mathcal{I} = \left((S_i)_{i=1}^N, \pi \right)$, where S_i is a measurable set of signals of player i, $S = \times_{i=1}^N S_i$ is the set of signal profiles, and $\pi \in \Delta(S \times \Theta)$ describes the joint distribution of signals and states.[3]

A *(Bayesian) game* is a pair $(\mathcal{G}, \mathcal{I})$. A *strategy* for player i in the game is a measurable mapping $\sigma_i : S_i \to \Delta(A_i)$. We write $\sigma_i(a_i|s_i)$ for the probability of an action a_i given the signal s_i. A strategy profile $\sigma = (\sigma_1, \ldots, \sigma_N)$ is associated with the product mapping

[3] We allow the information structure to be infinite, while the other objects in the model are finite. This richness is necessary to accommodate the full range of possible higher order beliefs and correspondingly the full range of equilibrium behavior across all counterfactuals.

$\sigma : S \to \Delta(A)$, where $\sigma(a|s) = \prod_{i=1}^{N} \sigma_i(a_i|s_i)$. Player i's expected utility under σ is[4]

$$U_i(\sigma) = \sum_{\theta \in \Theta} \int_{s \in S} \sum_{a \in A} u_i(a, \theta)\, \sigma(a|s)\, \pi(ds, \theta).$$

The profile σ is a *Bayes Nash equilibrium* if $U_i(\sigma) \geq U_i(\sigma_i', \sigma_{-i})$ for all i and for all strategies σ_i', and σ_{-i} refers to the strategy profile of all agents but i. (We shall frequently refer to Bayes Nash equilibrium simply as equilibrium.)

An *outcome* of $\mathcal{G}$ is a distribution $\phi \in \Delta(A \times \Theta)$. Note that ϕ contains all the information required in order to compute players' payoffs or any welfare criterion that only depends on realized actions and states. The outcome ϕ is *induced* by a pair $(\mathcal{I}, \sigma)$, where σ is a strategy in $(\mathcal{G}, \mathcal{I})$, if

$$\phi(a, \theta) = \int_{s \in S} \sigma(a|s)\, \pi(ds, \theta). \tag{1}$$

A *prior* is a distribution $\mu \in \Delta(\Theta)$. Given a prior μ and a base game $\mathcal{G}$, Bergemann and Morris (2013, 2016) defined a *Bayes correlated equilibrium (BCE) of* $(\mu, \mathcal{G})$ as an outcome $\phi(a, \theta)$ that satisfies the *obedience constraints*: for all i, a_i, and a_i',

$$\sum_{\theta \in \Theta} \sum_{a_{-i} \in A_{-i}} \left(u_i(a_i, a_{-i}, \theta) - u_i(a_i', a_{-i}, \theta)\right) \phi(a_i, a_{-i}, \theta) \geq 0, \tag{2}$$

where a_{-i} refers to the strategy profile of all agents but i, and the *Bayes consistency condition*: for all $\theta \in \mathit{\Theta}$

$$\sum_{a \in A} \phi(a, \theta) = \mu(\theta). \tag{3}$$

The obedience constraints require that whenever the joint distribution $\phi(a_i, a_{-i}, \theta)$ recommends action a_i to player i, then player i will find it in their interest to choose the recommended action a_i rather than any other action a_i'. The obedience condition is a best-response condition that is based on the information conveyed by the recommendation alone.

[4] Given a function $f : S \times \Theta \to \mathbb{R}$, the expression $\sum_{\theta \in \Theta} \int_{s \in S} f(s, \theta)\, \pi(ds, \theta)$ is the Lebesgue integral of f with respect to the measure that assigns a mass $\pi(X \times \{\theta\})$ for every $X \subseteq S$ and $\theta \in \Theta$. Thus, we emphasize the finiteness of Θ and the (possible) infiniteness of S.

The Bayes consistency condition states the marginal distribution of the outcome distribution over the state space equals the common prior distribution μ. The marginal (distribution) of θ is computed as usual from the joint distribution $\phi(a, \theta)$ by summing up over all of the values of action profiles a.

Theorem 1 in Bergemann and Morris (2016) establishes that ϕ is a BCE of $(\mu, \mathcal{G})$ if and only if there exists an information structure $\mathcal{I} = (S, \pi)$ and an equilibrium σ of $(\mathcal{G}, \mathcal{I})$ such that ϕ is induced by $(\mathcal{I}, \sigma)$ and μ is the marginal distribution of π on S.[5] In fact, Theorem 1, establishes a revelation principle familiar from the mechanism design literature. Namely, it is without loss of generality to restrict attention to information structures where the signal of each player can be identified with the action recommendation. Thus, there is a canonical class of information structures with which we can induce any outcome that can be induced with some information structure and equilibrium.

For the purpose of the current analysis, we shall use a weaker notion of BCE that only imposes the obedience conditions (2), and does not impose Bayes consistency (3) with respect to any particular prior. This formulation is convenient for applications in which the prior is unknown and has to be inferred from observable data. Formally, we say that the outcome ϕ is a *BCE of the base game* $\mathcal{G}$ if it is a BCE of $(\mu, \mathcal{G})$ for some prior μ. The corresponding and logically equivalent statement to Theorem 1 in Bergemann and Morris (2016) is that ϕ is a BCE of the base game $\mathcal{G}$ if and only if there exists an information structure $\mathcal{I}$ and an equilibrium σ of $(\mathcal{G}, \mathcal{I})$ such that ϕ is induced by $(\mathcal{I}, \sigma)$. Since the prior is described as part of the information structure, we will hereafter mostly avoid explicit reference to μ.

[5] Bergemann and Morris (2016) assume that S is finite, but the extension to arbitrary measurable S is routine.

3 Joint Predictions and Counterfactuals

Before developing our results on counterfactuals, we address the following question: Suppose that the players were to simultaneously play K games, $\mathcal{G}^1, ..., \mathcal{G}^K$, where $\mathcal{G}^k = \left(A_i^k, u_i^k\right)_{i=1}^N$. What joint predictions can we make about the outcomes that would arise in equilibria of the respective games, assuming that the players have the same information structure in every game? We have already observed that the set of outcomes that could arise in game $\mathcal{G}^k$ is the set of BCE. But because players have the same information in each of the K games, there will be additional consistency requirements across the games arising from fixed information. The purpose of this section is to characterize those extra restrictions. Our main result regarding counterfactual behavior and other later results will be established using this characterization when we impose particular structure or interpretations on the games $\mathcal{G}^k$.

3.1 Joint Predictions with Fixed Information

We say that an outcome profile

$$\left(\phi^1, ..., \phi^K\right) \in \Delta\left(A^1 \times \Theta\right) \times \cdots \times \Delta\left(A^K \times \Theta\right)$$

is a *joint prediction* if there exists an information structure $\mathcal{I}$ and, for each $k = 1, \ldots, K$, an equilibrium σ^k of $\left(\mathcal{G}^k, \mathcal{I}\right)$ such that ϕ^k is induced by $\left(\mathcal{I}, \sigma^k\right)$.

The *linked game* $\overline{\mathcal{G}} = \left(\overline{A}_i, \overline{u}_i\right)_{i=1}^N$ is defined by, for each i, $\overline{A}_i = A_i^1 \times \cdots \times A_i^K$ and

$$\overline{u}_i\left(\overline{a}, \theta\right) = \sum_{k=1,...,K} u_i^k\left(a^k, \theta\right),$$

where $\overline{a}_i = \left(a_i^1, ..., a_i^K\right)$. We refer to $\mathcal{G}^k$ as a *component game* of $\overline{\mathcal{G}}$. An outcome $\overline{\phi}$ of $\overline{\mathcal{G}}$ can be identified with a joint distribution in $\Delta\left(A^1 \times .... \times A^K \times \Theta\right)$.[6]

[6] The most closely related antecedent to linked games we know of is the symmetrization of two-player games due to Brown and von Neumann (1950). Their symmetric game is essentially a linked game whose component games are simply the original game and its "permutation," where the players trade roles.

Theorem 1 (Joint Predictions)**.**

A tuple $\left(\phi^1,...,\phi^K\right)$ *is a joint prediction for* $\mathcal{G}^1,...,\mathcal{G}^K$ *if and only if there exists a BCE* $\overline{\phi}$ *of* $\overline{\mathcal{G}}$ *for which the marginal distribution of* $\overline{\phi}$ *on* $A^k \times \Theta$ *is* ϕ^k *for each* $k = 1,\dots,K$*.*

Proof of Theorem 1. We first establish two preliminary claims. Fix an information structure $\mathcal{I}$ and strategy profile $\overline{\sigma}$ in $\left(\overline{\mathcal{G}},\mathcal{I}\right)$. For each k, let σ_i^k be the strategy in $\left(\mathcal{G}^k,\mathcal{I}\right)$ where $\sigma_i^k\left(\cdot|s_i\right)$ is the marginal of $\overline{\sigma}\left(\cdot|s_i\right)$ on A_i^k.

Claim 1: If $\overline{\phi}$ is the outcome induced by $\left(\mathcal{I},\overline{\sigma}\right)$ and ϕ^k is the outcome induced by $\left(\mathcal{I},\sigma^k\right)$, then ϕ^k is the marginal of $\overline{\phi}$ on $A^k\times\Theta$. This follows from

$$\begin{aligned}
\phi^k\left(a^k,\theta\right) &= \int_{s\in S}\sigma^k\left(a^k|s\right)\pi\left(ds,\theta\right)\\
&= \int_{s\in S}\sum_{a^{-k}\in A^{-k}}\overline{\sigma}\left(a^k,a^{-k}|s\right)\pi\left(ds,\theta\right)\\
&= \sum_{a^{-k}\in A^{-k}}\int_{s\in S}\overline{\sigma}\left(a^k,a^{-k}|s\right)\pi\left(ds,\theta\right)\\
&= \sum_{a^{-k}\in A^{-k}}\overline{\phi}\left(a^k,a^{-k},\theta\right),
\end{aligned}$$

where $A^{-k}=\prod_{k'\neq k}A^{k'}$.

Claim 2: $\overline{\sigma}$ is an equilibrium of $\left(\overline{\mathcal{G}},\mathcal{I}\right)$ if and only if σ^k is an equilibrium of $\left(\mathcal{G}^k,\mathcal{I}\right)$ for each k. This follows from the identity

$$\begin{aligned}
\overline{U}_i\left(\overline{\sigma}\right) &= \sum_{\theta\in\Theta}\int_{s\in S}\sum_{\overline{a}\in\overline{A}}\bar{u}_i\left(\overline{a},\theta\right)\overline{\sigma}\left(\overline{a}|s\right)\pi\left(ds,\theta\right)\\
&= \sum_{\theta\in\Theta}\int_{s\in S}\sum_{k=1,\dots,K}\left[\sum_{a\in A^k}u_i^k\left(a,\theta\right)\sigma^k\left(a|s\right)\right]\pi\left(ds,\theta\right)\\
&= \sum_{k=1,\dots,K}\sum_{\theta\in\Theta}\int_{s\in S}\left[\sum_{a\in A^k}u_i^k\left(a,\theta\right)\sigma^k\left(a|s\right)\right]\pi\left(ds,\theta\right)\\
&= \sum_{k=1,\dots,K}U_i^k\left(\sigma^k\right).
\end{aligned}$$

Now, if $\bar{\sigma}$ is not an equilibrium, then there exist i and a strategy $\bar{\tau}_i$ such that

$$\sum_{k=1,\dots,K} U_i^k\left(\sigma^k\right) = \overline{U}_i\left(\overline{\sigma}\right) < \overline{U}_i\left(\bar{\tau}_i, \overline{\sigma}_{-i}\right) = \sum_{k=1,\dots,K} U_i^k\left(\tau_i^k, \sigma_{-i}\right),$$

where τ_i^k is the marginal distribution of $\overline{\tau}_i$ on A_i^k. Thus, for at least one k, τ_i^k is a profitable deviation in $\left(\mathcal{G}^k, \mathcal{I}\right)$. Conversely, if there is a profitable deviation in one of the component games, say to τ_i^k for player i in $\mathcal{G}^k$, then the composite strategy $\overline{\tau}_i\left(\overline{a}_i|s_i\right) = \tau_i\left(\overline{a}_i^k|s_i\right)\sum_{\tilde{a}_i^k}\overline{\sigma}_i\left(\tilde{a}_i^k, \overline{a}_i^{-k}|s_i\right)$ is a profitable deviation in the linked game. This completes the proof of Claim 2.

With these claims in hand, we now prove the theorem. For the if direction, suppose that $\left(\phi^1, \dots, \phi^K\right)$ is a joint prediction. Then there exist an information structure $\mathcal{I}$ and, for each k, an equilibrium σ^k of $\left(\mathcal{G}^k, \mathcal{I}\right)$ such that ϕ^k is induced by $\left(\mathcal{I}, \sigma^k\right)$. Define strategies for the linked game $\overline{\sigma}_i\left(\overline{a}_i|s_i\right) = \prod_{k=1}^K \sigma_i^k\left(\overline{a}_i^k|s_i\right)$, and let $\overline{\phi}$ be the outcome induced by $\left(\mathcal{I}, \overline{\sigma}\right)$. By Claim 1, ϕ^k is the marginal of $\overline{\phi}$ on $A^k \times \Theta$. By Claim 2, $\overline{\sigma}$ is an equilibrium of $\left(\overline{\mathcal{G}}, \mathcal{I}\right)$ and hence, by Theorem 1 in Bergemann and Morris (2016), $\overline{\phi}$ is a BCE of $\overline{\mathcal{G}}$, as desired.

For the only if direction, suppose that $\overline{\phi}$ is a BCE of $\overline{\mathcal{G}}$ with marginals $\left(\phi^1, \dots, \phi^K\right)$. Then by Theorem 1 in Bergemann and Morris (2016), there exists an information structure $\mathcal{I}$ and equilibrium $\overline{\sigma}$ of $\left(\overline{\mathcal{G}}, \mathcal{I}\right)$ such that $\overline{\phi}$ is induced by $\left(\mathcal{I}, \overline{\sigma}\right)$. Now, for each i and k, define the strategy $\sigma_i^k\left(\cdot|s_i\right)$ to be the marginal of $\overline{\sigma}_i\left(\cdot|s_i\right)$ on A_i^k. Then for each k, by Claim 1, ϕ^k is induced by $\left(\mathcal{I}, \sigma^k\right)$, and by Claim 2, σ^k is an equilibrium of $\left(\mathcal{G}^k, \mathcal{I}\right)$. Hence, $\left(\phi^1, \dots, \phi^K\right)$ is a joint prediction, as desired. □

In effect, Theorem 1 shows that there is revelation principle for joint predictions. In particular, by extension from Theorem 1 in Bergemann and Morris (2016), there is a canonical class of information structures that suffice to rationalize any joint outcome of the component games. Namely, each player's signal is simply their profile of recommended actions for each component game. This is a natural extension of the logic behind BCE: A player must at least know which action they played in each component game, so a necessary condition for equilibrium is that the player not want to deviate conditional on all of their recommended actions. Moreover, it is sufficient for them to know just their recommended actions in order to implement equilibrium strategies that induce the observed outcomes.

Note that because of the additive separability of payoffs, the obedience constraints (2) specialized to the linked game reduce to the smaller set of constraints that, for all i, $\overline{a}_i \in \overline{A}_i$, k, and $a_i' \in A_i^k$,

$$\sum_{\theta \in \Theta} \sum_{\overline{a}_{-i} \in \overline{A}_{-i}} \left(u_i^k \left(\overline{a}_i^k, \overline{a}_{-i}^k, \theta \right) - u_i^k \left(a_i', \overline{a}_{-i}^k, \theta \right) \right) \overline{\phi} \left(\overline{a}_i, \overline{a}_{-i}, \theta \right) \geq 0. \tag{4}$$

Thus, an outcome of the linked game is obedient as long as each player does not benefit from deviating in any component game, conditional on their actions in all component games.

3.2 Counterfactuals when Information is Latent and Fixed

We now apply the preceding result to counterfactual predictions. An analyst has data about play in an *observed game* $\mathcal{G}$. The sets of possible states, actions, and players' payoff functions are known, but the analyst does not know the information structure $\mathcal{I}$ and may have limited data on the equilibrium distribution of states and actions.[7]

There is data on behavior in the observed game. We want to predict behavior in the unobserved game. The data is described by a set outcomes which are consistent with observation:

$$M \subseteq \Delta \left(A \times \Theta \right).$$

There are various possible specifications for M, which represent different kinds of observed data. In many cases, the set M can be taken to be a set of outcomes that satisfy moment equalities or inequalities. For example:

[7] The assumption that states, actions, and payoffs are known is without loss of generality. This is discussed in the Online Appendix. Our results also generalize to the case where there is more than one observed game and more than one counterfactual game, as discussed in Section 7.

1. $M = \{\phi\}$ for some particular ϕ. This corresponds to the case described in the introduction, where the joint distribution of states and actions is observed. Then the prior on states is known, but the rest of the information structure $\mathcal{I}$ is a latent parameter.

2. $M = \{\phi \in \Delta(A \times \Theta) \,|\mathrm{marg}_A \phi = \psi\}$ for some $\psi \in \Delta(A)$. In this case, the joint distribution of actions is known, but the information structure (including the distribution of θ) is latent.

3. $M = \{\phi \in \Delta(A \times \Theta) \,|\mathrm{marg}_A \phi \in \Psi\}$ for some $\Psi \subseteq \Delta(A)$. In this case, we do not even observe the entire distribution of the players' actions. For example, it could be that only some statistic, such as the average action or the highest action is observed.

More specifically, the analyst knows that the outcome ϕ of the observed game (i) lies in a set $M \subseteq \Delta(A \times \Theta)$ (where M is derived from the data), (ii) it was generated under some information structure $\mathcal{I}$, and (iii) it was induced by an equilibrium of $(\mathcal{G}, \mathcal{I})$. The analyst wants to make counterfactual predictions for what might happen if the *unobserved game* $\widehat{\mathcal{G}}$ were played. But in making the counterfactual prediction, the analyst wants to assume that the information structure $\mathcal{I}$ remains the same. We adopt the convention that unaccented objects correspond to the (partially) observed game and circumflex accented objects correspond to the unobserved game, e.g., outcomes for the two games are denoted ϕ and $\widehat{\phi}$, respectively.

We ask which outcomes $\widehat{\phi}$ could be induced by some equilibrium of $\left(\widehat{\mathcal{G}}, \mathcal{I}\right)$? Formally, an outcome $\widehat{\phi} \in \Delta\left(\widehat{A} \times \Theta\right)$ is a *counterfactual prediction* if there exist an information structure $\mathcal{I}$ and equilibria σ and $\widehat{\sigma}$ of $(\mathcal{G}, \mathcal{I})$ and $\left(\widehat{\mathcal{G}}, \mathcal{I}\right)$, respectively, such that the outcome ϕ induced by σ is in M and such that $\widehat{\phi}$ is induced by $\widehat{\sigma}$. The set of counterfactual predictions is denoted $\widehat{\Phi}$.

Our second main result is a characterization of $\widehat{\Phi}$ in terms of the linked game $\overline{\mathcal{G}}$ with component games $\mathcal{G}$ and $\widehat{\mathcal{G}}$. As in Theorem 1, it is the BCE characterization that will link the equilibrium behavior in the observed and the unobserved game.

Theorem 2 (Counterfactual Predictions)**.**
An outcome $\widehat{\phi} \in \Delta\left(\widehat{A} \times \Theta\right)$ *is in* $\widehat{\Phi}$ *if and only if there is a BCE* $\overline{\phi}$ *of* $\overline{\mathcal{G}}$ *such that (i) the marginal distribution of* $\overline{\phi}$ *on the observed game* $A \times \Theta$ *is in* M *and (ii)* $\widehat{\phi}$ *is the marginal distribution of* $\overline{\phi}$ *on the unobserved game* $\widehat{A} \times \Theta$*.*

Proof of Theorem 2. By the definition of a counterfactual prediction, $\widehat{\phi} \in \widehat{\Phi}$ if and only if there is a $\phi \in M$ such that $\left(\phi, \widehat{\phi}\right)$ is a joint prediction for $\overline{\mathcal{G}}$. By Theorem 1, such a joint prediction exists if and only if there exists a BCE $\overline{\phi}$ of the linked game $\overline{\mathcal{G}}$ for which ϕ and $\hat{\phi}$ are marginals on $A \times \Theta$ and $\hat{A} \times \Theta$, respectively, and $\phi \in M$, as desired. □

As discussed earlier, we can represent different kinds of data through the choice of M. Broadly, the set M can be used to represent various moment restrictions from the data, such as when only a subset of actions or statistics of actions can be observed. In each of these cases, M can be described as the intersection of a finite number of linear inequalities. As a result, $\widehat{\Phi}$ is also described by finitely many linear inequalities, being the projection onto $\Delta(A \times \Theta)$ of the set of BCE of the linked game which satisfy the finitely many obedience constraints and the constraints corresponding to M. In the special case of counterfactuals, the obedience constraints for the linked game (4) reduce to, for all i, a_i, $\widehat{a}_i$, a_i', and $\widehat{a}_i'$,

$$\begin{aligned}
&\sum_{\theta \in \Theta} \sum_{a_{-i} \in A_{-i}} \sum_{\widehat{a}_{-i} \in \widehat{A}_{-i}} \left(u_i\left(a_i, a_{-i}, \theta\right) - u_i\left(a_i', a_{-i}, \theta\right)\right) \overline{\phi}\left(\left(a_i, \widehat{a}_i\right), \left(a_{-i}, \widehat{a}_{-i}\right), \theta\right) \geq 0; \\
&\sum_{\theta \in \Theta} \sum_{a_{-i} \in A_{-i}} \sum_{\widehat{a}_{-i} \in \widehat{A}_{-i}} \left(\widehat{u}_i\left(\widehat{a}_i, \widehat{a}_{-i}, \theta\right) - \widehat{u}_i\left(\widehat{a}_i', \widehat{a}_{-i}, \theta\right)\right) \overline{\phi}\left(\left(a_i, \widehat{a}_i\right), \left(a_{-i}, \widehat{a}_{-i}\right), \theta\right) \geq 0.
\end{aligned} \tag{5}$$

If we fix a Bayesian welfare criterion $w(\widehat{a}, \theta)$ over ex post counterfactual outcomes, then the range of expected values of w across all counterfactuals can be obtained by solving a pair of finite-dimensional linear programs. We give specific examples of these linear programs in the next section.

a/θ	-1	1
0	0	0
1	$-1+z$	$1+z$

Table 1: Payoffs as a function of actions and states.

4 Two Examples

We now illustrate the content of Theorem 2 using two examples. The first is a discrete version of the Roy (1951) model of self selection. The second is an entry game in the spirit of Ciliberto and Tamer (2009) and Magnolfi and Roncoroni (2021). These examples demonstrate how the obedience constraints and moment conditions obtained from the data can be used to make counterfactual predictions without explicitly identifying the latent prior and/or information structure. We will also explore how the counterfactual prediction depends on what we observe about the outcome of the observed game, as captured by the moment conditions.

4.1 One-Player Games and the Roy Model

We specialize to the case where $N = 1$, in which case the "game" is a decision problem being solved by a single decision maker. We further specialize to the case where there are two states $\Theta = \{-1, 1\}$ and two actions $A = \{0, 1\}$.[8] In the observed game, the player's payoff is $u(a, \theta) = a\theta$. Thus, we have normalized the payoff from $a = 0$ to zero, which is without loss of generality as long as the optimal action depends on the state. In the counterfactual, the payoff is $\widehat{u}(a, \theta) = a(\theta + z)$ for some $z \in \mathbb{R}$, so that $a = 1$ has become more or less valuable in both states by the same amount. The observed game corresponds to $z = 0$:

This problem can be viewed as a special and discretized version of the canonical model of self-selection, first formulated by Roy (1951), in which the player decides whether or not to opt in to some activity. Opting out (action 0) results in a certain payoff (normalized

[8] We will note in the text how some results generalize beyond binary actions/states, and report these results as propositions in the Supplementary Appendix.

a/θ	-1	1
0	α	$\frac{1}{2}-\alpha$
1	$\frac{1}{2}-\alpha$	α

Table 2: Observed distribution of actions and states.

to 0), but opting in (action 1) yields an uncertain payoff.[9] The player may have imperfect knowledge of the state that informs their decision of whether to opt in. The Roy model has been described by Heckman (2010b) as "the prototype for many models of self-selection in economics" (p. 264).[10] In Roy (1951), opting in represents a decision to choose one occupation over another; in Willis and Rosen (1979) or Carneiro et al. (2003) opting in represents one schooling level over another. In the context of employment, the payoff represents the difference in the player's potential long-run earnings between the two occupations. In the experimental setting, the opt in payoff is the average potential treatment effect, which may be uncertain due to latent characteristics of the player that affect the treatment's efficacy. In the counterfactual, opting in becomes more or less valuable, which corresponds to a shift in the distribution of potential wages or the distribution of potential treatment effects.

Counterfactual Welfare with Observable Outcomes We will first consider the case where the analyst observes the entire distribution of actions and states. This corresponds to the case originally considered by Roy (1951), in which it is assumed that wages are observed for whatever occupation is chosen. The specific observed outcome (a, θ) is given by the probabilities in Table 2 for some $\alpha \in [1/4, 1/2]$.[11] Thus, both states are equally likely and the probability that the player chooses the ex post optimal action is the same in both states.

[9] The constant payoff associated with opting out is a notable feature—simplifying our exposition—whereas the Roy (1951) model and most of the subsequent work has random payoffs associated with either choice of action.

[10] See Heckman and Vytlacil (2007) for a canonical exposition. Abbring and Heckman (2007) survey a recent literature that develops econometric methods for identifying the player's information, which we briefly discuss in Section 6.

[11] The constraint $\alpha \leq 1/2$ is necessary for the probability of every outcome to be non-negative. If $\alpha < 1/4$, the player's payoff would be negative, which is inconsistent with utility maximization, since the player can always opt out and obtain a payoff of zero.

We first investigate the maximal counterfactual welfare that is consistent with the outcome in the observed game. The maximal counterfactual welfare is the value of the following linear program:

$$\max_{\overline{\phi}\geq 0} \sum_{(a,\widehat{a},\theta)} \overline{\phi}\,(a,\widehat{a},\theta)\,\widehat{a}\,(\theta+z)$$

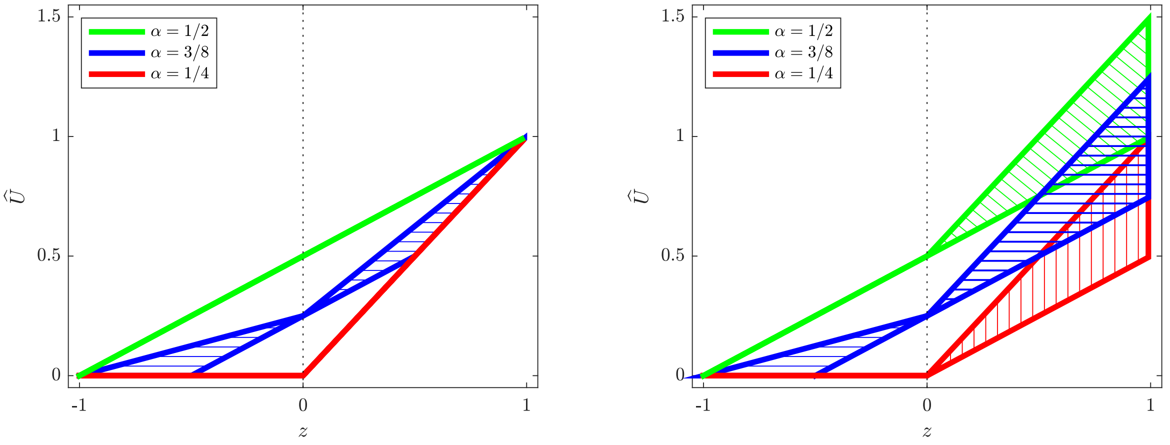
$$\text{s.t.}\ \sum_{\widehat{a}} \overline{\phi}\,(a,\widehat{a},\theta) = \begin{cases} \alpha, & \text{if } (a,\theta)\in\{(0,-1),(1,1)\}; \\ \frac{1}{2}-\alpha, & \text{otherwise;} \end{cases} \tag{6a}$$

$$\sum_{\theta} \overline{\phi}\,(0,\widehat{a},\theta)\,\theta \leq 0 \quad \text{and} \quad \sum_{\theta} \overline{\phi}\,(1,\widehat{a},\theta)\,\theta \geq 0,\ \forall \widehat{a}; \tag{6b}$$

$$\sum_{\theta} \overline{\phi}\,(a,0,\theta)\,(\theta+z) \leq 0 \quad \text{and} \quad \sum_{\theta} \overline{\phi}\,(a,1,\theta)\,(\theta+z) \geq 0,\ \forall a. \tag{6c}$$

The objective is counterfactual welfare. The first constraints given by (6a) requires that the marginal on the observed game is the observed outcome. These equality constraints are the generalizations of the Bayes consistency conditions. The first set of inequality constraints given by (6b) are the obedience conditions for the observed game. The second set of inequality constraints given by (6c) are the obedience conditions for the counterfactual game. The program that establishes the minimum counterfactual welfare is the same, except that we now minimize the objective rather than maximize.

Counterfactual welfare is plotted in the left panel of Figure 1 for $\alpha \in \{0.25, 0.375, 0.5\}$. If $\alpha = 0.5$, then the player opts in if and only if the state is 1. This is possible only if the player has full information about the state. If we change z, the player will continue to make the ex post optimal choice, so that counterfactual welfare is the green line. If $\alpha = 0.25$, then the player enters with the same probability independent of the state. This is possible only if the player has no information: Otherwise there would be some correlation between the action and the state. Again, in the counterfactual, the player still has no information, and the counterfactual payoff is given by the red line. Thus, for $\alpha \in \{0.25, 0.5\}$ we get a point prediction for counterfactual welfare. In these two extremes of the outcome



(a) Fully observed outcome

(b) Partially observed outcome

Figure 1: Counterfactual welfare in the Roy model.

distribution parametrized by α, the outcome in the observed game allows the analyst to completely deduce the distribution of interim beliefs held by the player. In consequence, the counterfactual welfare prediction in the one agent environment is unique.

If $\alpha = 0.375$, then the player's action is ex post optimal less often than can be explained by full information and more often than can be explained by no information. Here are two information structures that rationalize the data, formally described as $\pi_1(s, \theta)$ and $\pi_2(s, \theta)$ below in Table 3:

(i) Half of the time, the signal $s \in \{-1, 1\}$ is fully informative and matches the state, and the player opts in if and only if the state is high. The other half of the time, the signal, $s = 0$, is uninformative, and the player opts in with probability 1/2.

(ii) The player's signal $s \in \{-1, 1\}$ is a noisy observation of the state and the probability that $s = \theta$ is 3/4. The player opts in when the signal is 1 and opts out when the signal is -1.

Suppose that the player's information is given by (i). As we vary z, the counterfactual payoff is simply the average of the full- and no-information payoffs. This is the top blue line in the left panel of Figure 1. In contrast, if information is given by (ii), then the player's

θ/s	-1	0	1
-1	$1/4$	$1/4$	0
1	0	$1/4$	$1/4$

θ/s	-1	1
-1	$3/8$	$1/8$
1	$1/8$	$3/8$

Table 3: Signal distributions $\pi_1(s, \theta)$ and $\pi_2(s, \theta)$

strategy when $z = 0$ remains optimal as long as $z \in [-1/2, 1/2]$. When $z < 1/2$, the noisy signal is not strong enough to induce the player to opt in, and it is optimal to always opt out. When $z > 1/2$, it is optimal to always opt in. The resulting welfare is given by the lower blue line in the left panel of Figure 1.

It turns out that the information structures (i) and (ii) achieve maximum and minimum counterfactual welfare. Thus, all we can say about the player's counterfactual welfare is that it lies in the hatched area between the blue lines. With regard to minimum counterfactual welfare, observe that whatever information the player has, it must be Blackwell more informative than the signal generated by the action the player chose, which corresponds precisely to the information structure (ii). Thus, welfare under (ii) must be weakly lower than welfare under any other information structure that rationalizes the player's behavior. With regard to maximum counterfactual welfare, consider what distributions of the player's belief about the state are consistent with the observed behavior. The obedience constraint implies that when $a = 0$ ($a = 1$), the player's belief that the state is high is less than $1/2$ (greater than $1/2$). In addition, the *average* belief is $1/4$ conditional on $a = 0$ and $3/4$ conditional on $a = 1$. Among distributions of beliefs satisfying these property, the one that is most informative in the sense of Blackwell (1953)—and is therefore associated with the highest payoff—assigns probability $1/4$ to a belief of 0, $1/4$ to a belief of 1, and $1/2$ to a belief of $1/2$. This is precisely the belief distribution induced by the information structure (i).

Note that for any value of α, there is a unique "local" counterfactual payoff prediction, where by local we mean that the counterfactual game is the same as the observed game. Thus, when $z = 0$, counterfactual welfare must be equal to observed welfare. The reason is that the player can always simply repeat their observed action in the counterfactual and achieve as high a payoff as in the observed outcome. Similarly, in the observed game, the player could have instead mimicked their action in the counterfactual and achieved as high a payoff as in the counterfactual outcome. Thus, the two payoffs must be equal.

Thus, we have shown that (1) minimum counterfactual welfare is attained when the player has only the information revealed by observed behavior; (2) maximum counterfactual welfare is attained when the player has the Blackwell most informative information structure consistent with observed behavior; and (3) local counterfactual welfare is unique. In the Online Appendix, we show that these properties generalize to any number of actions, and (1) and (3) also generalize to any number of states.[12] All three results rely on the fact that there is a single player and we are looking at counterfactuals about that player's payoff. In the Online Appendix, we give an example of a single-player game where there is not a unique counterfactual when we look at behavior instead of payoffs. We also see the failure of the uniqueness of the local counterfactual in the entry game in the next section. However, the local counterfactual for welfare is unique in two-player zero-sum games, as is shown in the Online Appendix.

Partially Observed Outcomes In randomized control trials, it is natural to suppose that the data is censored, and we only observe average treatment effects for players who opt into the trial. Similarly, if "opting in" represents a decision to join the labor force, we would likely not observe potential income for the unemployed. We now revisit the analysis of the previous section, but supposing that we only observe the distribution of the state conditional on $a = 1$. In the linear program, we simply replace the first constraint with

$$\sum_{\widehat{a},\theta} \overline{\phi}\left(0, \widehat{a}, \theta\right) = 1/2, \ \sum_{\widehat{a}} \overline{\phi}\left(1, \widehat{a}, -1\right) = \frac{1}{2} - \alpha, \text{ and } \sum_{\widehat{a}} \overline{\phi}\left(1, \widehat{a}, 1\right) = \alpha.$$

[12] With more than two states, there may not be a Blackwell most informative information structure consistent with observed behavior.

Counterfactual welfare is plotted in the right panel of Figure 1. For $z < 0$, the prediction is unchanged. The reason is that for these counterfactuals, opting in is less attractive than in the observed game, so whenever we observed the player opt out, they will also opt out in the counterfactual, in which case the payoff is independent of the state.

When $z > 0$, however, the player may opt in when we observed them opt out. Exactly how often this happens and what welfare results depends on the state distribution that rationalizes the data. Consider first when $\alpha = 0.5$, i.e., we observe the player opt in half the time, but when the player opts in, the state is always high. We do not know the distribution of the state when the player opts out. In the case considered previously, the state was always low when the player opted out. This is the most pessimistic case, and corresponds to the lower green curve in the right panel of Figure 1. However, obedience only requires that the state be low at least half the time when the player opts out; in particular, the data can also be rationalized by the state being equally likely to be high or low when the player opted out. In this case, the counterfactual payoff would be strictly higher, since when $z > 0$, the player would strictly prefer to opt in, thereby achieving the welfare on the higher green line. Any payoff between the green lines can also be rationalized. A similar analysis explains the upper and lower bounds for $\alpha = 0.375$ and $\alpha = 0.25$. Interestingly, the direction in which our bounds expand depends on the particular observed outcome: When the observed welfare is high, it is the upper bound on counterfactual welfare that is relaxed, whereas when observed welfare was low, it is the lower bound that is relaxed. In the intermediate case, both upper and lower bounds on counterfactual welfare are relaxed.

4.2 An Entry Game

Our second application is a simple entry game with entry costs, in the spirit of Ciliberto and Tamer (2009) and Magnolfi and Roncoroni (2021). Two firms choose whether to enter (E) or not enter (N) a market. The payoff from not entering is zero. Each firm $i = 1, 2$ has a cost to enter the market $c_i \in \{0, 2\}$. If a single firm enters, that firm earns a monopoly revenue

3. If both firms enter, they each earn revenue 1. Payoffs are summarized in the following matrix:

$$\begin{array}{ccc}
a_1/a_2 & N & E \\
\text{N} & (0,0) & (0,3-c_2) \\
\text{E} & (3-c_1,0) & (1-c_1,1-c_2)
\end{array}$$

To map this game into our framework, we identify the state with the ordered pair of firms' entry costs, that is, $\theta = (c_1, c_2)$.

The outcome of the observed game is as follows: First, all entry-cost profiles are equally likely. Second, each firm enters if and only if their cost is low. This outcome corresponds to the unique symmetric equilibrium if each firm only knows their own cost. We assume that the analyst observes this outcome. In the counterfactual prediction, we compute producer surplus $\widehat{U}$ when we add z to the payoff from entry. This can be interpreted as a change in firms' entry costs. In Figure 2, we have plotted the counterfactual predictions for $\widehat{U}$ for z between -3 and 2.

A detailed analysis is given in Section 8.2 in the Online Appendix, where we also explicitly write out the linear programs corresponding to maximum and minimum $\widehat{U}$. We now give a brief summary. First consider which information structures are consistent with the observed outcome. Since each firm's action is perfectly correlated with their cost, and since firms know their own actions, they must know their own costs as well. In addition, when $z = 0$, entering is strictly dominant when the cost is low. Thus, the data does not place any further restrictions on firms' information about the other firm's cost when their own cost is low. However, the data does restrict the beliefs of high-cost firms. In particular, in the observed outcome, each firm would be just indifferent between entering and not entering if they had no information about the other firm's cost. If a high-cost firm's signal contained non-trivial information about the other firm's cost, then they would sometimes believe that the other firm is more likely to enter than under the prior, in which case it would be strictly optimal to enter, thus violating obedience.

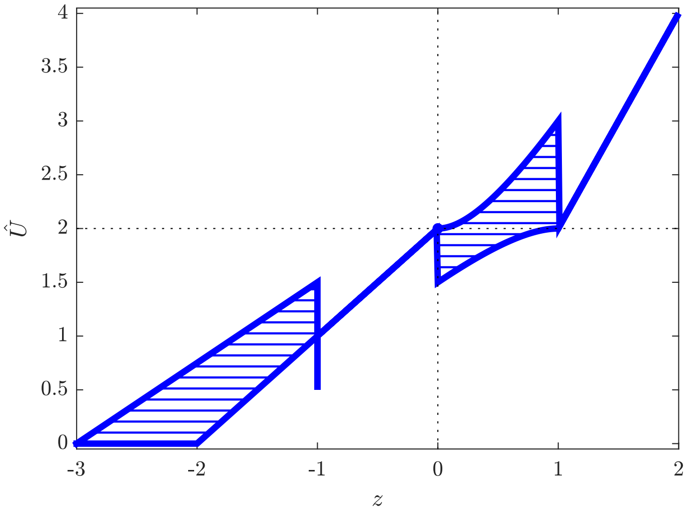


Figure 2: Counterfactual welfare in the entry game.

With this characterization in hand, we can now explicitly describe the information structures and equilibria that attain the maximum and minimum counterfactual predictions. As indicated in Figure 2, there are several cases, depending on the value of z. This discussion will focus on $z > -1$. The remaining cases are described in the Online Appendix.

First, if $z > 1$, then entering is strictly dominant regardless of the entry cost, so that there is a point prediction for welfare. There is also a unique counterfactual prediction for aggregate payoffs for $z \in (-1, 0)$, which is attained in the equilibrium in which firms enter if and only if their costs are low.

The analysis for $z \in [0, 1]$ is more subtle. In this range, there is a symmetric mixed-strategy equilibrium in which high-cost firms independently randomize between entering and not entering. Since a high-cost firm's payoff is zero (the payoff from not entering), all surplus is generated by the low-cost firms. Moreover, a low-cost firm's payoff from entering is just the high-cost firm's payoff from entering plus 2 (the difference in entry cost), so that aggregate payoff is 2 in this equilibrium. However, we can see in Figure 2 that counterfactual welfare can be both higher and lower than in the symmetric mixed equilibrium. This is possible if firms' entry decisions are correlated.

In particular, suppose that in addition to learning their entry costs, firms observe the outcome of a randomization device which recommends either enter ($s_i' = e$) or not enter ($s_i' = n$) with probabilities:

$$\begin{array}{ccc} s_1'/s_2' & n & e \\ n & 1-\beta-2\gamma & \gamma \\ e & \gamma & \beta \end{array}$$

where $\beta \geq 0$, $\gamma \geq 0$, and $\beta + 2\gamma \leq 1$, and (s_1', s_2') is independent of (c_1, c_2). There are choices of β and γ for which it is an equilibrium for low-cost firms to ignore the correlation device and always enter and for high-cost firms to enter if and only if $s_i' = e$. To maximize producer surplus, we choose the parameters of the correlation device to maximize the probability of one firm entering, subject to only one firm entering at a time. The latter condition corresponds to $\beta = 0$, in which case the obedience constraint for a high-cost firm not entering reduces to $\gamma \leq z$. Setting γ equal to this upper bound, we obtain the highest counterfactual producer surplus. (In the Online Appendix, we verify that the obedience constraint for high-cost firms not entering is also satisfied.)

Similarly, to minimize producer surplus, we choose the parameters to maximize the probability that two firms enter by setting $\beta = 1 - 2\gamma$, in which case the obedience constraint for not entering reduces to $\gamma \leq (1-z)\,/\,(2-z)$. Setting γ equal to its upper bound, we obtain the lowest counterfactual producer surplus. When $z = 0$, this outcome involves exactly one high-cost firm entering. This lowers the aggregate payoff below the equilibrium in which high-cost firms do not enter, because it lowers the payoffs of low-cost firms. This illustrates the multiplicity of local counterfactual predictions when there is more than one player, due to multiplicity of equilibria; it is also an equilibrium for firms to ignore the correlation device and play the symmetric mixed-strategy equilibrium.

5 Tightening Counterfactuals and Varying Information

We now describe some straightforward extensions of the model that are relevant for applications. These extensions could be used to either shrink or expand the set of counterfactual predictions obtained from Theorem 2. At a conceptual level, one can conceive of a wide variety of refinements and relaxations of our notion of a counterfactual prediction, where we impose different assumptions on players' information and strategies and how they can vary across games. A key question is whether these variations will retain the tractable linear structure that we characterized for our baseline model. The particular extensions that we propose are notable in that the resulting set of counterfactuals is still given by the intersection of finitely many linear inequalities, which correspond to obedience constraints and data restrictions. Since the extensions are rather notationally burdensome, we provide an informal discussion, but there is a more complete and formal treatment in the working paper version, see Bergemann, Brooks, and Morris (2020).

5.1 Tightening Counterfactuals

Bounds on Information First, we can enrich the model with further restrictions on players' information. The most straightforward such restriction is to require that the information structure that rationalizes the data be at least as informative as some fixed information structure $\widetilde{\mathcal{I}} = \left(\left\{\widetilde{S}_i\right\}_{i=1}^{N}, \widetilde{\pi}\right)$ in the sense of *individual sufficiency* (Bergemann and Morris, 2016). This means that players observe signals in $\widetilde{\mathcal{I}}$ (or an equivalent information structure), but may have access to additional information. To incorporate this assumption, we expand the outcome to a joint distribution over $(a, \theta, s) \in A \times \Theta \times \widetilde{S}$, and we write $\phi(a, \theta, \widetilde{s})$, etc. The moment restriction M can now incorporate assumptions about the joint distribution of $(\widetilde{s}, \theta)$. We also need to modify the obedience constraints (4). Since player i can observe $\widetilde{s}_i$, they can condition on it when they evaluate a deviation. Hence, the obedience constraints for the linked game are now that for every i, $\widetilde{s}_i$, a_i, $\widehat{a}_i$, a'_i, and $\widehat{a}'_i$,

$$\sum_{\theta\in\Theta}\sum_{a_{-i}\in A_{-i}}\sum_{\widehat{a}_{-i}\in\widehat{A}_{-i}}\sum_{\widetilde{s}_{-i}\in S_{-i}}\left(u_i\left(a_i,a_{-i},\theta\right)-u_i\left(a'_i,a_{-i},\theta\right)\right)\overline{\phi}\left(\left(a_i,\widehat{a}_i\right),\left(a_{-i},\widehat{a}_{-i}\right),\theta,\widetilde{s}_i,\widetilde{s}_{-i}\right)\geq 0;$$
$$\sum_{\theta\in\Theta}\sum_{a_{-i}\in A_{-i}}\sum_{\widehat{a}_{-i}\in\widehat{A}_{-i}}\sum_{\widetilde{s}_{-i}\in S_{-i}}\left(\widehat{u}_i\left(\widehat{a}_i,\widehat{a}_{-i},\theta\right)-\widehat{u}_i\left(\widehat{a}'_i,\widehat{a}_{-i},\theta\right)\right)\overline{\phi}\left(\left(a_i,\widehat{a}_i\right),\left(a_{-i},\widehat{a}_{-i}\right),\theta,\widetilde{s}_i,\widetilde{s}_{-i}\right)\geq 0. \tag{7}$$

The analogue of Theorem 2 with the lower bound on information is as follows: An outcome $\widehat{\phi}\in\Delta\left(\widehat{A}\times\Theta\times\widetilde{S}\right)$ is a counterfactual prediction if and only if there exists a BCE $\overline{\phi}\in\Delta\left(\overline{A}\times\Theta\times\widetilde{S}\right)$, satisfying the generalized obedience conditions (7), such that $\widehat{\phi}$ is the marginal of $\overline{\phi}$ on $\widehat{A}\times\Theta\times\widetilde{S}$, and the marginal of $\overline{\phi}$ on $A\times\Theta\times\widetilde{S}$ is in M. When $\widetilde{\mathcal{I}}$ is uninformative (i.e., $\widetilde{s}$ is independent of θ), this characterization reduces to Theorem 2. But as $\widetilde{\mathcal{I}}$ becomes more informative in the sense of individual sufficiency, the set of counterfactual predictions must shrink. (This follows immediately from Theorem 2 of Bergemann and Morris (2016).)

A particularly relevant special case is that of *public regressor*s, i.e., covariates x that are common knowledge among the players when they choose their actions. In this case, we can set $\widetilde{s}_i = x$ for all i, and $\tilde{\pi}$ captures the correlation between x and θ. This is how public regressors are modeled in Magnolfi and Roncoroni (2021).[13] Another important case is *private values*: In many applications, such as auctions, the state can be decomposed as $\theta=(\theta_1,\ldots,\theta_N)$ into player-specific components, and player i's preferences depend only on θ_i. In addition, we often assume that each player knows their own payoff type, which we can model by setting $\widetilde{s}_i=\theta_i$. For example, in the entry example of Section 4.2, the components of θ are the firms' costs, and we could impose a lower bound on information that each firm knows their own cost (as is generally assumed in the applied literature). In fact, this

[13] A conceptually distinct way to model the public regressor is by regarding each realization of x as a separate instance of the observed and counterfactual games, and by forming a larger linked game consisting of component games $\left(\mathcal{G}_x,\widehat{\mathcal{G}}_x\right)_{x\in X}$. This formulation implicitly assumes that the players' signals are conditionally independent of x, whereas the approach developed in the main text allows for correlation between signals and public regressors.

lower bound would not have changed the counterfactual prediction in our example, since the observed outcome revealed that firms knew their costs anyway. However, if we observed a different outcome, say, that from the symmetric mixed-strategy equilibrium where the benefits from entering would be sufficiently larger than 2, then the data would be consistent with firms not knowing their own costs, and assuming private costs would lead to a tighter counterfactual prediction. (Incidentally, Magnolfi and Roncoroni (2021) assume that entry costs are private, and Syrgkanis et al. (2021) consider private-value auctions in which each bidder knows their own value for the good being sold.)

Note that the observed outcome is itself a lower bound on information, in the sense that players must know the action they took in the observed game. Similarly, the observed outcome is also an upper bound on information: Players cannot have so much information that they would have preferred to deviate from their observed actions. In many cases, this upper bound on information is abstract and difficult to interpret. But we can contemplate games for which there is a natural interpretation of the implied upper bounds on information. Consider a game in which the players do not interact at all; they simply try to guess the state, and are incentivized to be as accurate as possible. If we were to observe players not always guessing the state correctly, then we would know that their information about the state must be bounded away from full information.[14]

Payoff Shifters A common device in applied microeconomics is to assume that there are exogenous "payoff shifters" that affect a player's utility and are observable to the player and to the econometrician (Tamer, 2003; Jia, 2008; Bajari et al., 2010; Somaini, 2020). Actions and shifters together provide much richer information about players' information than just the action. Such payoff shifters are easily incorporated into our framework.

[14] An alternative approach would be to restrict attention to information structures that are less informative than some fixed $\widetilde{\mathcal{I}}$. But this is equivalent to requiring that there exist kernels $K_i : \widetilde{S}_i \to \Delta(A_i)$ such that

$$\sum_{\overline{a}_{-i} \in \overline{A}_{-i}} \overline{\phi}\left(\overline{a}_i, \overline{a}_{-i}, \widetilde{s}_i, \widetilde{s}_{-i}, \theta\right) = K_i\left(\overline{a}_i | \widetilde{s}_i\right) \sum_{\overline{a}' \in \overline{A}} \overline{\phi}\left(\overline{a}', \widetilde{s}_i, \widetilde{s}_{-i}, \theta\right), \tag{8}$$

a constraint that is non-linear in K_i and $\overline{\phi}$.

For each i, let ω_i be a shifter for player i's payoffs in the observed game, which takes values in a finite set Ω_i. Player $i's$ payoff is now of the form $u_i(a, \theta, \omega_i)$. Suppose ω_i is directly observable to player i but not to players $-i$ and let $\eta(\omega|\theta)$ denote the conditional distribution of the entire profile of payoff shifters. (We are assuming that η can be separately identified from the data.) In addition, we assume that players do not learn anything about ω from their signals in the information structure. Since each player can observe their component of ω, they may condition on it when choosing their action. We can therefore reduce this game to a normal form in which player i's action is a strategy ξ_i that maps payoff shifters into a pure action a_i. (It is without loss to restrict player i to pure mappings, since we allow them to mix over pure strategies, as in the baseline model.) Player i's payoff as a function of the strategy profile is

$$u(\xi, \theta) = \sum_{\omega \in \Omega} \eta(\omega|\theta) \, u_i(\xi(\omega), \theta, \omega).$$

We can now immediately apply Theorem 2 to the normal form game in which a player's action is their strategy.

The value of the payoff shifters is that the implied richness of the observed game may reveal more information about players' information. We can illustrate this with a variation on the Roy model of Section 4.1. Suppose we first make the counterfactual problem harder, by having the analyst not observe anything about the state in the original game. However, we suppose that the player observes a payoff shifter ω, which is distributed uniformly on $[-1, 1]$ and independent of the player's signal and the state.[15] The payoff from opting out is still normalized to zero, but the payoff to a player who opts in is $\theta + 1 - 2\omega$.[16]

[15] We let the payoff shifter be a continuous variable. It would be straightforward but slightly less elegant to make the same points with a finite set of realizations (consistent with the rest of our model), in which case we would be able to identify the cumulative distribution of the player's belief at finitely many points.

[16] This functional form is chosen so that we obtain a clean expression for the distribution of beliefs. Note that this model is neither a generalization nor a special case of the model of Section 4.1.

Further suppose that we observe the joint distribution of (a, ω). Let p denote the interim probability that the player assigns to $\theta = 1$. Opting out is optimal only if

$$p - (1 - p) + 1 - 2\omega \leq 0 \iff p \leq \omega.$$

Thus, the probability of opting out given ω is precisely the probability that p is less than ω. Indeed, if we write $P(\omega)$ for the probability of $a = 0$ given ω, then $P(\omega)$ is the cumulative distribution of the player's interim belief. This distribution of the interim beliefs describes all features of the player's information that are relevant for counterfactual predictions. In that sense, the payoff shifter allows us to point identify the information structure.

5.2 Variable Information

Our focus in this paper is on counterfactuals when the base game changes but fundamentals and information are held fixed. This notion of counterfactual is entirely standard: We change one feature of the environment (the base game) and keep everything else the same. It is also implicitly adopted in the vast majority of applied work on incomplete information games, generally along with even stronger functional form assumptions (e.g., independent or affiliated private values in the literature on auctions). At the same time, while fixed information is a natural benchmark, we may also wish to consider cases where information can vary between the observed game and the counterfactual. This would be reasonable if the players have opportunities to acquire more or different information between the observed game and the counterfactual. We now describe a range of counterfactual exercises, where the prior over states is held fixed, but players' information about the state is allowed to vary between the observed and counterfactual games. In fact, we can accommodate a variety of such "variable-information" counterfactuals within our framework, and different assumptions about how information can vary can be captured by suitably modifying the obedience constraints (5).[17]

In the most extreme case, one could allow information to vary in an arbitrary way between the observed game and the counterfactual. This corresponds to the unrestricted-information counterfactuals approach discussed in the introduction and applied by Magnolfi and Roncoroni (2021) and Syrgkanis et al. (2021). In this case, we simply have to modify the obedience constraints for the linked game: When deciding whether or not to deviate in the observed game (resp. counterfactual game), players should only condition on their action in the observed game (resp. counterfactual game), since the action in the other game may be based on entirely different information. Thus, the obedience constraints become, for all i, a_i, a_i', $\widehat{a}_i$, and $\widehat{a}_i'$:

$$\sum_{\theta\in\Theta}\sum_{a_{-i}\in A_{-i}}\sum_{\widehat{a}\in\widehat{A}}\left(u_i\left(a_i,a_{-i},\theta\right)-u_i\left(a_i',a_{-i},\theta\right)\right)\overline{\phi}\left(a_i,a_{-i},\widehat{a},\theta\right)\geq 0; \tag{9a}$$

$$\sum_{\theta\in\Theta}\sum_{\widehat{a}_{-i}\in\widehat{A}_{-i}}\sum_{a\in A}\left(\widehat{u}_i\left(\widehat{a}_i,\widehat{a}_{-i},\theta\right)-\widehat{u}_i\left(\widehat{a}_i',\widehat{a}_{-i},\theta\right)\right)\overline{\phi}\left(a,\widehat{a}_i,\widehat{a}_{-i},\theta\right)\geq 0. \tag{9b}$$

This is equivalent to allowing arbitrary BCE in the observed and counterfactual games, with the only requirement being that the BCE have the same marginal distribution over the fundamental.

Alternatively, one could assume that in the counterfactual game, players know as much as they knew in the observed game, but they may know more. In this case, the obedience constraints for observed actions (9a) stay the same, but the constraints for the counterfactual

[17] For many of the same reasons as with variable information, we may also wish to consider cases where the distribution of fundamentals varies as we vary the base game. Indeed, we view the fixed-information hypothesis in much the same way as we view fixed fundamentals: a natural benchmark, but one that should be regarded with appropriate skepticism, depending on the application. It would be straightforward to adapt our methodology to allow for some features of the state distribution to vary between observation and counterfactual. Of course, if the base game, information, and fundamentals can all change in an unrestricted manner, then connection between the observed and unobserved games is lost, and our so-called counterfactual prediction would simply reduce to BCE of the unobserved game.

(9b) become, for all i, a_i, $\widehat{a}_i$, and $\widehat{a}_i'$:

$$\sum_{\widehat{a}_{-i},a_{-i},\theta} \left(\widehat{u}_i\left(\widehat{a}_i,\widehat{a}_{-i},\theta\right)-\widehat{u}_i\left(\widehat{a}_i',\widehat{a}_{-i},\theta\right)\right)\overline{\phi}\left(a_i,a_{-i},\widehat{a}_i,\widehat{a}_{-i},\theta\right)\geq 0.$$

Thus, when deciding whether or not to deviate in the counterfactual game, players are able to condition on both their observed and counterfactual actions.

Going a step further, we could even combine this approach with the lower bounds on information discussed in Section 5.1, and suppose that in the counterfactual game, the players observe signals $\widetilde{s}$ (with known distribution) which were not available in the observed game.[18] In this case, the obedience constraints for the counterfactual (9b) become, for all i, a_i, $\widetilde{s}_i$, $\widehat{a}_i$, and $\widehat{a}_i'$:[19]

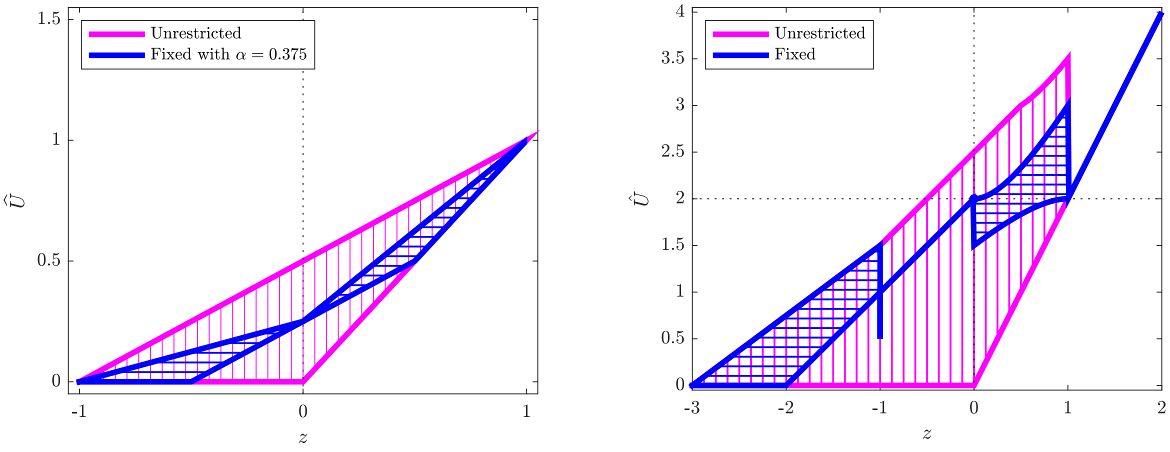
$$\sum_{\widehat{a}_{-i},a_{-i},\widetilde{s}_{-i},\theta} \left(\widehat{u}_i\left(\widehat{a}_i,\widehat{a}_{-i},\theta\right)-\widehat{u}_i\left(\widehat{a}_i',\widehat{a}_{-i},\theta\right)\right)\overline{\phi}\left(a_i,a_{-i},\widehat{a}_i,\widehat{a}_{-i},\widetilde{s},\theta\right)\geq 0.$$

This is by no means an exhaustive list: The methodology in this paper can accommodate a wide variety of assumptions about how information varies between observed and counterfactual games. One can also allow for players to have less information in the counterfactual, which would be relevant, say, in a counterfactual where we impose that a monopolist cannot price based on some observable characteristic of consumers.

As an illustration of unrestricted-information counterfactuals, we computed unrestricted-information counterfactuals in the examples of Section 4, which are depicted in Figure 3. As expected, the counterfactual prediction is more permissive with unrestricted information than with fixed information. For the Roy model, the only restriction from the data is that both states are equally likely. Blackwell's theorem (Blackwell, 1953) implies that counterfactual

[18] Dickstein and Morales (2018) consider such a counterfactual in the context of international trade, with a more parametric model of firms' information when they decide whether or not to export.

[19] An important subtlety in this counterfactual is that players observe $\widetilde{s}$ *and possibly more*. The reason is that when we use $\tilde{a}_i$ to represent player i's information in the counterfactual game, we cannot separate the "part" of $\tilde{a}_i$ that depends on $\widetilde{s}$ from that which depends on information available in the observed game. Imposing the restriction that players observe a new orthogonal signal $\widetilde{s}$ and nothing more would require us to explicitly model the information that gave rise to observed behavior.



(a) Roy model

(b) Entry game

Figure 3: Counterfactuals with fixed vs unrestricted information.

welfare is maximized when the player has full information, in which case the player chooses the optimal ex post action. On the other hand, counterfactual welfare is minimized when they have no information, in which case opting in is optimal only if $z \geq 0$. The counterfactual predictions for welfare is everything between these two extremes. There is no more that we can say.

6 Related Empirical and Econometric Work

A growing number of papers has already successfully employed the BCE characterization in developing unrestricted-information counterfactuals; notably (as discussed in the introduction) Magnolfi and Roncoroni (2021) and Syrgkanis, Tamer, and Ziani (2021).[20] We now describe these papers in more detail, and explain their connection to our results.

Magnolfi and Roncoroni (2021) consider an entry game with binary actions. Firms' preferences and the distribution of entry costs depend on a parameter $\omega \in \Omega$.[21] Firms' entry costs correspond to the payoff-relevant state θ in the present paper. Magnolfi and

[20] Both Magnolfi and Roncoroni (2021) and Syrgkanis, Tamer, and Ziani (2021) were circulated before the present paper. We cite only the most recent versions.

[21] Magnolfi and Roncoroni (2021) denote the parameter by θ. We use the letter ω to avoid confusion with the notation used in this paper for the payoff-relevant state.

Roncoroni (2021) suppose that firms know their own entry costs but may have arbitrary information about other firms' entry costs, which can be identified with BCE in which firms know their own costs (cf. the discussion in Section 5.1). Magnolfi and Roncoroni (2021) identify the subset of Ω for which there exists such a BCE that matches the observed distribution of actions. Thus, the presumption is that the parameter ω is the same for all realizations of the game that contributed to the data. In their main counterfactual analysis, the base game changes, the parameter ω is held fixed, but the information can vary in an arbitrary manner between the observed and counterfactual games. This corresponds to the unrestricted-information counterfactual that we discussed in Section 5.2, which is more permissive than the fixed-information counterfactual characterized by Theorem 2. Following earlier versions of this paper, Magnolfi and Roncoroni (2021) added an analysis of fixed-information counterfactuals and found that fixed information substantially reduces the range of counterfactual outcomes.

Syrgkanis, Tamer, and Ziani (2021) develop a theory and applications of what we call unrestricted-information counterfactuals. In contrast to Magnolfi and Roncoroni (2021), Syrgkanis, Tamer, and Ziani (2021) do not parametrize players' preferences or the distribution of the payoff-relevant state. Rather, they implicitly describe the identified set of state distributions as marginals of the BCE that are consistent with the observed distribution of actions. In this respect, their analysis makes a more direct use of the BCE characterization than Magnolfi and Roncoroni (2021), who use random set theory and support function to characterize the identified set of parameter values. Syrgkanis, Tamer, and Ziani (2021) use their results to analyze bidding behavior in generalized second-price and first-price auctions. Syrgkanis, Tamer, and Ziani (2021) compute counterfactual predictions when the auction format changes, say, from a generalized second-price auction to a first-price auction, and when information can vary arbitrarily between the observed and counterfactual games. We suspect that their empirical approach can be extended to fixed-information counterfactuals.

The scope of our paper is limited to the question of partial identification of counterfactuals, when the "data" is in the form of an exact moment (or moments) of the population outcome. We do not address the fundamental issue of how to conduct inference on the counterfactual prediction when one only has access to a noisy estimate derived from a finite data set, which of course plays a central role in the empirical papers described above.

In related work, Gualdani and Sinha (2020) maintain the fixed counterfactuals assumption in their application to a single agent discrete choice model of voting. Canen and Song (2020) introduce a decomposition approach used frequently in labor economics to offer counterfactual predictions in strategic settings using BCE as the solution concept.

A common and novel feature of all of this work is a general and non-parametric treatment of the information structure. By contrast, the vast literature on the Roy model, our leading example in Section 4, has addressed imperfect information largely with a parametric approach, see Abbring and Heckman (2007) for a comprehensive survey. As emphasized by Heckman (2010a), agents will base their treatment decisions on their beliefs about the outcome, referred to as *ex-ante evaluations* or *ex-ante outcomes*, whereas the econometrician typically observes the *ex-post outcomes*. A recent strand of literature, including Carneiro et al. (2003), Cunha et al. (2005) among others, has developed methods for identifying agents' information and separating the ex-ante and ex-post outcome distributions. The identification of the information structure is typically based on the assumption that the returns from treatment depends in a known manner on a specific set of factors, and the agent's information is identified with the subset of factors that are observed when treatment is chosen.

Our approach models information entirely non-parametrically and treats it as a nuisance parameter that is of interest only insofar as it influences the counterfactual prediction. Building on the prior results on informationally-robust identification and inference with unrestricted information due to Magnolfi and Roncoroni (2021) and Syrgkanis, Tamer, and Ziani (2021), we established that we can both allow for arbitrary and unobservable information structures and yet hold the information structure fixed when computing counterfactuals.

This insight could be used in a wide range of microeconometric applications, including those mentioned above.

7 Conclusion

Our first result (Theorem 1) characterized the precise implications of Bayesian rationality and common priors for joint predictions in games, under the hypothesis of fixed information. Theorem 2 applied that result to derive counterfactual predictions under fixed and latent information. We have shown that there is a sharp description of the set of counterfactual outcomes that are consistent with observed data. The main virtues of our approach are its analytical tractability and the minimalist assumptions about the form of information and equilibrium selection. These two go hand in hand: it is precisely because we allow a large set of information structures and equilibria that the set of counterfactual predictions has a simple linear structure. In spite of the weakness of our assumptions, we have demonstrated through examples that the predictive power of fixed information can be significant, in particular compared to what can be predicted if we do not fix information between observation and counterfactual. Moreover, we have shown that it is possible to refine the counterfactual predictions with further assumptions on the information structure, although only certain kinds of assumptions about information will preserve the linear structure.

Theorem 1 has implications beyond counterfactuals, and we will conclude the paper by discussing three. First, Theorem 1 implies a simple test of the hypothesis that the players had the same information in the component games; observed outcomes can be rationalized by a common information structure if and only if there exists a BCE of the linked game for which the marginals are the observed outcomes.

Second, Theorem 1 can be used to generate *informationally-robust rankings* of games, as we now explain. Suppose that there are two games of interest $\mathcal{G}^k$ for $k = 1, 2$. An analyst assigns a value to each outcome of $\mathcal{G}^k$. This value is used to compare outcomes across games.

For example, $\mathcal{G}^1$ and $\mathcal{G}^2$ may be entry games with different barriers to entry, and the value is social surplus. Or they may represent different auction formats and the value is expected revenue. The question is, which game is associated with a higher value?

Let us say that $\mathcal{G}^1$ *dominates* $\mathcal{G}^2$ if for every information structure $\mathcal{I}$, and equilibria of $(\mathcal{G}^1,\mathcal{I})$ and $(\mathcal{G}^2,\mathcal{I})$, the induced outcome for $\mathcal{G}^1$ has higher value than to the induced outcome of $\mathcal{G}^2$.[22] Theorem 1 immediately implies the following: $\mathcal{G}^1$ dominates $\mathcal{G}^2$ if and only if for every BCE $\overline{\phi}$ of the linked game with component games $\mathcal{G}^1$ and $\mathcal{G}^2$, the value of ϕ^1 is greater than that of ϕ^2, where ϕ^k is the marginal of $\overline{\phi}$ on $\mathcal{G}^k$. When the games are finite, determining dominance with fixed information reduces to checking the feasibility of a finite system of linear inequalities.

Third, we could combine the informationally-robust rankings and counterfactual predictions. In particular, we could suppose that there is an observed game $\mathcal{G}$ and two counterfactual games $\widehat{\mathcal{G}}^1$ and $\widehat{\mathcal{G}}^2$. We can ask: does $\widehat{\mathcal{G}}^1$ dominate $\widehat{\mathcal{G}}^2$, subject to observed behavior in $\mathcal{G}$ and the constraint that the information structure is the same in all three games? One can answer this question by first forming the linked game with components $\mathcal{G}$, $\widehat{\mathcal{G}}^1$, and $\widehat{\mathcal{G}}^2$. Then, $\widehat{\mathcal{G}}^1$ dominates $\widehat{\mathcal{G}}^2$ if and only if for every BCE of the linked game that is consistent with the data, the marginal on $\widehat{\mathcal{G}}^1$ has greater value than the marginal on $\widehat{\mathcal{G}}^2$. In the Online Appendix, we report examples of such informationally-robust rankings on counterfactuals for both the Roy model and the entry game. Further developing these and other applications of Theorem 1 is an important direction for future work.

[22] We thank Jeff Ely for suggesting this notion of dominance.

References

Abbring, J. and J. Heckman (2007): *Econometric Evaluation of Social Programs, Part III: Distributional Treatment Effects, Dynamic Tretment Effects, Dynamic Discrete Choice, and General Equilibrium Policy Evaluation*, North-Holland, vol. 6B of *Handbook of Econometrics*, chap. 72, 5145–5303.

Anscombe, F. and R. Aumann (1963): "A Definition of Subjective Probability," *Annals of Mathematical Statistics*, 34, 199–205.

Bajari, P., H. Hong, and S. P. Ryan (2010): "Identification and Estimation of a Discrete Game of Complete Information," *Econometrica*, 78, 1529–1568.

Bergemann, D., B. Brooks, and S. Morris (2017): "First Price Auctions with General Information Structures: Implications for Bidding and Revenue," *Econometrica*, 85, 107–143.

——— (2020): "Counterfactuals with Latent Information," Working paper, Yale University, University of Chicago, and Massachusetts Institute of Technology.

——— (2021): "Counterfactuals with Latent Information," Tech. Rep. 2162R2, Cowles Foundation Discussion Paper 2162R2.

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

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

Blackwell, D. (1953): "Equivalent Comparison of Experiments," *Annals of Mathematics and Statistics*, 24, 265–272.

Brown, G. and J. von Neumann (1950): "Solutions of Games by Differential Equations," *Contributions to the Theory of Games*, 73.

Canen, N. and K. Song (2020): "A Decomposition Approach to Counterfactual Analysis in Game-Theoretic Models," Tech. rep.

Carneiro, P., K. Hansen, and J. Heckman (2003): "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice," *International Economic Review*, 44, 361–422.

Ciliberto, F. and E. Tamer (2009): "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, 77, 1791–1828.

Cunha, F., J. Heckman, and S. Navarro (2005): "Separating Uncertainty from Heterogeneity in Life Cycle Earnings," *Oxford Economic Papers*, 57, 191–262.

Dickstein, M. and E. Morales (2018): "What Do Exporters Know?" *Quarterly Journal of Economics*, 133, 1753–1801.

Gualdani, C. and S. Sinha (2020): "Identification and Inference in Discrete Choice Models with Imperfect Information," Tech. rep.

Heckman, J. (2010a): "Building Bridges between Structural and Program Evaluation Approaches in Evaluating Policy," *Journal of Economic Literature*, 48, 356–398.

Heckman, J. and E. Vytlacil (2007): *Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation*, North-Holland, vol. 6B of *Handbook of Econometrics*, chap. 70, 4779–4874.

Heckman, J. J. (2010b): "Selection Bias and Self-Selection," in *Microeconometrics*, ed. by S. N. Durlauf and L. E. Blume, Springer, 242–266.

Heumann, T. (2019): "Informationally Robust Comparative Statics in Incomplete Information Games," Tech. rep.

Jia, P. (2008): "What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retailing Industry," *Econometrica*, 76, 1263–1316.

Magnolfi, L. and C. Roncoroni (2021): “Estimation of Discrete Games with Weak Assumptions on Information,” Tech. rep.

Peski, M. (2008): “Comparison of Information Structures in Zero-Sum Games,” *Games and Economic Behavior*, 62, 732–735.

Roy, A. (1951): “Some Thoughts on the Distribution of Earnings,” *Oxford Economic Papers*, 3, 135–146.

Savage, L. (1954): *The Foundations of Statistics*, New York: Wiley, 1st ed.

Somaini, P. (2020): “Identification in Auction Models with Interdependent Costs,” *Journal of Political Economy*, 128, 3820–3871.

Syrgkanis, V., E. Tamer, and J. Ziani (2021): “Inference on Auctions with Weak Assumptions on Information,” Tech. rep.

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

Willis, R. and S. Rosen (1979): “Education and Self-Selection,” *Journal of Political Economy*, 87, 7–36.

8 Online Appendix

This appendix is organized as follows. Section 8.1 presents general results on single-player counterfactuals and contains additional examples based on the Roy model of Section 4, including informationally-robust rankings as described in Section 7. Section 8.2 provides a more formal and complete analysis of the entry game of Section 4.2. Section 8.3 studies counterfactual predictions in two-player zero-sum games. Section 8.4 studies counterfactuals in a first-price auction with reserve price. Section 8.5 discusses how various nominal assumptions in the model of Section 2 are in fact normalizations and are without loss of generality.

8.1 Single Player Example: Further Analysis

In Section 4.1, we explained how in single-player games, minimum counterfactual welfare is obtained with the minimally informative information structure, in which a player's signal is their observed action. We now give a general statement of this result:

Proposition 1 (Minimum single-player counterfactual welfare)**.**
Suppose $N = 1$, and fix an observed decision problem $\mathcal{G} = (A, u)$ and moment restriction $M = \{\phi\}$. Define an information structure $\mathcal{I} = (S, \pi)$ by $S = A$ and such that $\pi(a, \theta) = \phi(a, \theta)$ for all a and θ. Then the obedient strategies are an equilibrium of $(\mathcal{G}, \mathcal{I})$, and $(\mathcal{I}, \sigma)$ induce the observed outcome. Moreover, for every counterfactual decision problem $\widehat{\mathcal{G}} = \left(\widehat{A}, \widehat{u}\right)$, the minimum expected counterfactual welfare across all counterfactual predictions is attained when the information structure is $\mathcal{I}$, and minimum counterfactual welfare is

$$\sum_{a \in A} \max_{\widehat{a} \in \widehat{A}} \sum_{\theta \in \Theta} \phi(a, \theta)\, \widehat{u}(\widehat{a}, \theta).$$

The proof is elementary, and follows the argument given in the text.

We next give a general statement of the result that with binary states, there is a maximally informative information structure which attains maximum counterfactual welfare.

When $\Theta = \{\theta_1, \theta_2\}$, we can represent the player's belief conditional on their signal as the probability that the state is θ_1. For each observed action $a \in A$, there is an interval of beliefs for which that action is optimal, which we can denote by $[x_L(a), x_H(a)]$. Conditional on taking the action a, every realized belief must be in this interval. The Blackwell-most informative belief distribution consistent with the data must have all of the mass concentrated on the end points of this interval. Any information structure that generates this distribution of beliefs will maximize the player's welfare in all counterfactual decision problems. One such information structure is as follows: Let $\mathcal{I} = (S, \pi)$ where $S = A \times \{L, H\}$, and define the conditional probabilities $\pi(a, H, \theta)$ and $\pi(a, L, \theta)$ to solve the following system of equations:

$$\begin{aligned}
\pi(a, H, \theta_1) + \pi(a, L, \theta_1) &= \phi(a, \theta_1);\\
\pi(a, H, \theta_2) + \pi(a, L, \theta_2) &= \phi(a, \theta_2);\\
\frac{\pi(a, H, \theta_1)}{\pi(a, H, \theta_1) + \pi(a, H, \theta_2)} &= x_H(a);\\
\frac{\pi(a, L, \theta_1)}{\pi(a, L, \theta_1) + \pi(a, L, \theta_2)} &= x_L(a).
\end{aligned}$$

When $x_L(a) < x_H(a)$, there is a unique solution:

$$\begin{aligned}
\pi(a, H, \theta_2) &= \frac{1 - x_H(a)}{x_H(a)} \frac{\phi(a, \theta_1) - \frac{x_L(a)}{1 - x_L(a)} \phi(a, \theta_2)}{1 - \frac{x_L(a)}{1 - x_L(a)} \frac{1 - x_H(a)}{x_H(a)}};\\
\pi(a, H, \theta_1) &= \frac{\phi(a, \theta_1) - \frac{x_L(a)}{1 - x_L(a)} \phi(a, \theta_2)}{1 - \frac{x_L(a)}{1 - x_L(a)} \frac{1 - x_H(a)}{x_H(a)}},
\end{aligned}$$

and $\pi(a, L, \theta) = \phi(a, \theta) - \pi(a, H, \theta)$. Otherwise, if $x^L(a) = x^H(a)$ (so that there is a unique belief at which a is a best response, which must be the belief conditional on being recommended a) then there is a continuum of solutions to this system, where $\pi(a, H, \theta_1) = \pi(a, H, \theta_2)$. Thus, we can just take $\pi(a, H, \theta_1) = \phi(a, \theta_1)$ and $\pi(a, H, \theta_2) = \phi(a, H, \theta_2)$. With this information structure, the player has an optimal strategy to choose a after the signals (a, H) and (a, L). Moreover, $\mathcal{I}$ is Blackwell-more informative than any other information structure that rationalizes the data. We have proven the following proposition:

Proposition 2.

Suppose that $N = 1$ and $\Theta = \{\theta_1, \theta_2\}$, and fix an observed game $\mathcal{G} = (A, u)$ and moment restriction $M = \{\phi\}$. Let the information structure $\mathcal{I}$ be as constructed in the preceding paragraph. Then the obedient strategies are an equilibrium of $(\mathcal{G}, \mathcal{I})$. Moreover, for every counterfactual decision problem $\widehat{\mathcal{G}} = \left(\widehat{A}, \widehat{u}\right)$, the maximum expected counterfactual welfare across all counterfactual predictions is attained when the information structure is $\mathcal{I}$, and maximum counterfactual welfare is

$$\sum_{(a,k)\in A\times\{L,H\}} \max_{\widehat{a}\in\widehat{A}} \sum_{\theta\in\Theta} \widehat{u}\left(\widehat{a}, \theta\right) \pi\left(a, k, \theta\right).$$

At a high level, this result depends on the fact that the set of distributions over beliefs partially ordered by mean-preserving spreads is a lattice when $|\Theta| = 2$. When $|\Theta| > 2$, this partially ordered set is no longer a lattice, and in particular, there need not be a most informative distribution of beliefs that rationalizes the data.

Finally, we argue that there is always a unique local counterfactual in single-player games:

Proposition 3.

Suppose that $N = 1$ and $M = \{\phi\}$. If the counterfactual game $\widehat{\mathcal{G}}$ is equal to $\mathcal{G}$, then there is a unique counterfactual welfare in all counterfactual predictions, which is welfare under ϕ:

$$\sum_{a\in A}\sum_{\theta\in\Theta} u\left(a, \theta\right) \phi\left(a, \theta\right).$$

The argument is that given in the text: Fix an information structure $\mathcal{I}$ and observed and counterfactual equilibrium strategies σ and $\widehat{\sigma}$ (that is, σ and $\widehat{\sigma}$ are optimal decision rules).

Since the two games are the same, the payoffs in the games are respectively

$$U = \sum_{\theta \in \Theta} \int_{s \in S} \sum_{a \in A} u(a, \theta) \sigma(a|s) \pi(ds, \theta) \text{ and } \widehat{U} = \sum_{\theta \in \Theta} \int_{s \in S} \sum_{a \in A} u(a, \theta) \widehat{\sigma}(a|s) \pi(ds, \theta).$$

But $\widehat{\sigma}$ is a feasible strategy in the observed game, so the fact that σ is an equilibrium must be $U \geq \widehat{U}$. By an analogous argument, $\widehat{U} \geq U$, so in fact they are equal. Finally, by the definition of a counterfactual prediction, we must have that σ and $\mathcal{I}$ induce ϕ, so that U is equal to welfare under ϕ.

Ranking two counterfactuals We will next use the Roy model to illustrate the methodology for ranking games discussed in Section 7. Specifically, we ask for which pairs of counterfactual parameters z^1 and z^2 does the agent always attain higher welfare under z^1 than under z^2, when we restrict attention to those information structures which are consistent with the observed outcome. Figure 4 plots the set of pairs $\left(\widehat{U}^1, \widehat{U}^2\right)$ of agent welfare obtained for the pairs $(z^1, z^2) = (0.5, 0.75)$ when the observed outcome corresponds to $\alpha = 0.375$, both under the assumption of fully-observable outcomes (the blue set) and partially-observable outcomes (the red set).

The picture clearly shows that agent is unambiguously better off under z^2. This can be seen from the fact that all of the sets lie above the 45 degree line. Indeed, this conclusion is theoretically trivial: the counterfactual with z^2 has payoffs that are pointwise higher, so that the agent could achieve a higher payoff with z^2 than with z^1 simply by using whatever strategy was optimal for z^1. Note that while this conclusion is theoretically obvious, it is not apparent in Figure 1: For many pairs $z^2 > z^1$, the set of possible welfare outcomes for the agent overlap. It is only by plotting agent welfare resulting from joint counterfactual predictions that we can see that higher values of z dominate.

Nonetheless, this example illustrates the power of fixing information when computing informationally-robust rankings: Without holding information fixed, there would be no dominance ranking between z^1 and z^2, whenever the two are sufficiently close together.

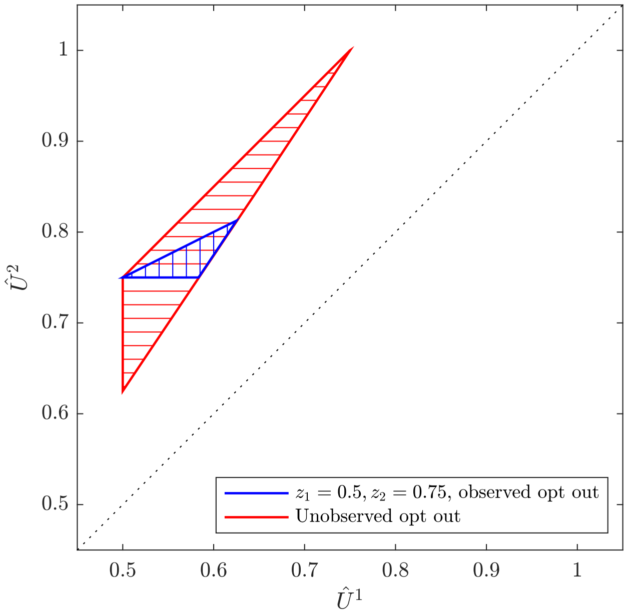


Figure 4: Ranking counterfactuals in the Roy model.

Welfare versus behavior In Section 4.1, we primarily focused on the player's welfare. This is not the only counterfactual outcome of interest. More broadly, we may ask how the player's *behavior* could change in the counterfactual, i.e., the probability of opting in for each state. While we do not analyze this question in detail, we can say that there are generally much weaker restrictions on behavior than on welfare. This is illustrated in Figure 5, which depicts the total probability that the player opts in as we vary z, for the cases considered above.

The left panel describes the counterfactual probability of opting in when we observe the entire outcome, including the state distribution when the player opts out. When the observed outcome is consistent with either either no information (the green curve) or full information (the blue curve), there is generically a point prediction for counterfactual behavior. However, for no information and $z = 0$, there are counterfactual predictions consistent with any opt-in probability between zero and one. This is true even though there is a point prediction for counterfactual welfare, simply because when $z = 0$, the player is indifferent between actions.

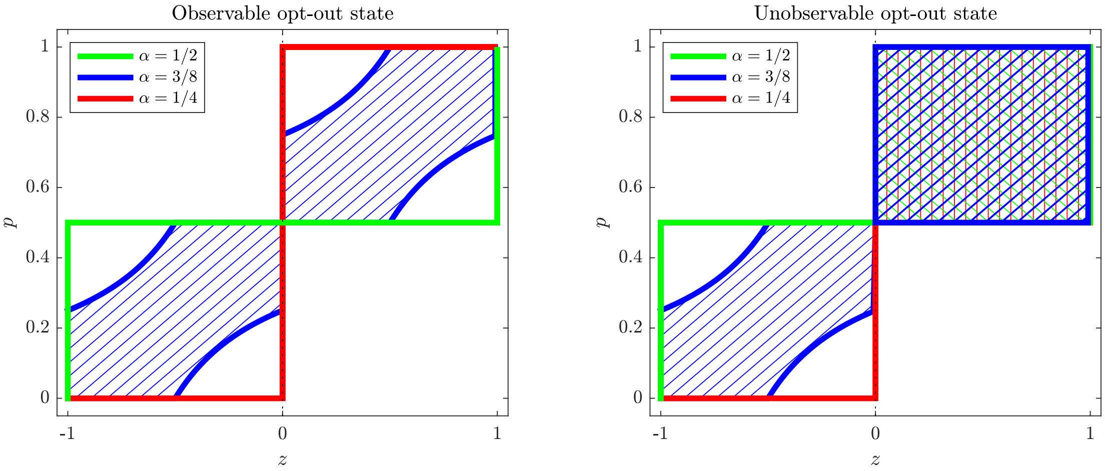


Figure 5: Counterfactual probability p of choosing the action $a = 1$.

For the intermediate case of partial information, there is always a fat set of counterfactual opt-in probabilities. Again, this is true even when $z = 0$, when there is a point prediction for welfare.

The counterfactual prediction for behavior when we do not observe the state after opting out is depicted in the right panel of Figure 5. The prediction is even more permissive in this case: For every $z > 0$, any opt-in probability between 1/2 and 1 is consistent with all three cases considered. For, in each of these examples, the player must always opt in when the state is good, and there is a state distribution that rationalizes the player's observed decision to opt out when $z = 0$ but such that they would strictly prefer to enter if $z > 0$.

8.2 Entry Game: Further Analysis

Linear program The linear program for maximum counterfactual producer surplus is

$$\max_{\overline{\phi}\geq 0} \sum_{(a,\widehat{a},c)} \left(\overline{\phi}\left(a,(E,N),c\right)(3-c_1)+\overline{\phi}\left(a,(N,E),c\right)(3-c_2)+\overline{\phi}\left(a,(E,E),c\right)(2-c_1-c_2)\right)$$

$$\text{s.t. } \sum_{\widehat{a}} \overline{\phi}\left(a,\widehat{a},c\right) = \begin{cases} \frac{1}{4} & \text{if } (a_i,c_i)\in\{(E,0),(N,2)\} \ \forall i; \\ 0 & \text{otherwise;} \end{cases}$$

$$\sum_{(\widehat{a}_2,c_2)} \left[\overline{\phi}\left(N,E,\widehat{a}_1,\widehat{a}_2,c_1,c_2\right)(3-c_1)+\overline{\phi}\left(N,N,\widehat{a}_1,\widehat{a}_2,c_1,c_2\right)(1-c_1)\right]\leq 0 \ \forall\left(\widehat{a}_1,c_1\right);$$

$$\sum_{(\widehat{a}_2,c_2)} \left[\overline{\phi}\left(E,E,\widehat{a}_1,\widehat{a}_2,c_1,c_2\right)(3-c_1)+\overline{\phi}\left(E,N,\widehat{a}_1,\widehat{a}_2,c_1,c_2\right)(1-c_1)\right]\geq 0 \ \forall\left(\widehat{a}_1,c_1\right);$$

$$\sum_{(\widehat{a}_1,c_1)} \left[\overline{\phi}\left(E,N,\widehat{a}_1,\widehat{a}_2,c_1,c_2\right)(3-c_2)+\overline{\phi}\left(N,N,\widehat{a}_1,\widehat{a}_2,c_1,c_2\right)(1-c_2)\right]\leq 0 \ \forall\left(\widehat{a}_2,c_2\right);$$

$$\sum_{(\widehat{a}_1,c_1)} \left[\overline{\phi}\left(E,E,\widehat{a}_1,\widehat{a}_2,c_1,c_2\right)(3-c_2)+\overline{\phi}\left(N,E,\widehat{a}_1,\widehat{a}_2,c_1,c_2\right)(1-c_2)\right]\geq 0 \ \forall\left(\widehat{a}_2,c_2\right);$$

$$\sum_{(a_2,c_2)} \left[\overline{\phi}\left(a_1,a_2,N,E,c_1,c_2\right)(3+z-c_1)+\overline{\phi}\left(a_1,a_2,N,N,c_1,c_2\right)(1+z-c_1)\right]\leq 0 \ \forall\left(a_1,c_1\right);$$

$$\sum_{(a_2,c_2)} \left[\overline{\phi}\left(a_1,a_2,E,E,c_1,c_2\right)(3+z-c_1)+\overline{\phi}\left(a_1,a_2,E,N,c_1,c_2\right)(1+z-c_1)\right]\geq 0 \ \forall\left(a_1,c_1\right);$$

$$\sum_{(a_1,c_1)} \left[\overline{\phi}\left(a_1,a_2,N,E,c_1,c_2\right)(3+z-c_2)+\overline{\phi}\left(a_1,a_2,N,N,c_1,c_2\right)(1+z-c_2)\right]\leq 0 \ \forall\left(a_2,c_2\right);$$

$$\sum_{(a_1,c_1)} \left[\overline{\phi}\left(a_1,a_2,E,E,c_1,c_2\right)(3+z-c_2)+\overline{\phi}\left(a_1,a_2,N,E,c_1,c_2\right)(1+z-c_2)\right]\geq 0 \ \forall\left(a_2,c_2\right).$$

The program for minimizing counterfactual producer surplus is the same, except that we change the maximization to minimization.

Detailed calculations for entry counterfactuals We analytically construct the equilibria that attain the boundaries of the numerically computed counterfactual prediction in Figure 2. We do not give a proof that these bounds are optimal.

Both firms always entering is an equilibrium if $z \geq 1$, and the resulting payoff is $2\,(1+z) - 2 = 2\,(3+z) - 6$. This is the unique counterfactual prediction when $z > 1$, when entering becomes strictly dominant.

When $z < 1$, always entering is not an equilibrium. As long as $z \geq 0$, there is a mixed strategy equilibrium in which low-cost firms always enter and a firm with high cost enters with probability α, to make the other firm indifferent between entering and not entering:

$$3 + z - (1+\alpha)/2 - 2 = 0 \iff \alpha = z.$$

Thus, these strategies are an equilibrium for $z \in [0,1]$. Since this equilibrium makes high-cost firms indifferent between entering and not entering, the payoff of the high-cost firm is zero, and the payoff when the cost is low is just the high cost, which is 2, so that the overall payoff in this equilibrium is 2.

We now construct equilibria for $z \in [0,1]$ that attain the upper and lower bounds of the counterfactual welfare. Firms observe the outcome of a correlation device that produces signals (s_1, s_2) that are independent of the firms' costs and has the following probabilities:

s_1/s_2	0	1
0	$1-\beta-2\gamma$	γ
1	γ	β

where $\gamma \in [0, 1/2]$ and $\beta \in [0, (1-\gamma)/2]$. In the equilibria we now construct, low-cost firms ignore this signal and always enter, but a high-cost firm i enters if and only if $s_i = 1$.

The obedience constraints are as follows: Conditional on $s_i = 1$, the likelihood of the other firm entering is $(\gamma + 2\beta)/(2\gamma + 2\beta)$. The reason is that the other firm will enter regardless of their signal if their cost is low, but will only enter if they get the high signal when their cost is high. Conditional on this signal, the payoff from entering must be non-negative:

$$1 + z - 2\frac{\gamma + 2\beta}{2(\gamma+\beta)} \geq 0.$$

Similarly, conditional on being told to not enter and having a high cost, the payoff from entering must be non-positive:

$$1 + z - 2\frac{1 - \beta - 2\gamma + 2\gamma}{2(1 - \beta - \gamma)} \leq 0.$$

The equilibrium payoffs are

$$\frac{1}{2}\left(3 + z - 2\frac{1 + \gamma + \beta}{2}\right) + \frac{1}{2}\left[(\gamma + \beta)(1 + z) - 2\frac{\gamma + 2\beta}{2}\right].$$

To obtain minimum counterfactual welfare, we set $\beta = 1 - 2\gamma$ and make the obedience constraint for entering hold as an equality. Intuitively, we are pushing down welfare by having firms enter with high probability. Solving for β, we obtain

$$\beta = 1 - 2\gamma = \frac{z}{2 - z}.$$

It is straightforward to verify that the obedience constraint for entering is always satisfied with these values for β and γ and $z \in [0, 1]$. The resulting aggregate payoff is

$$2 + z - \frac{1}{2 - z}.$$

which coincides with the simulated minimum counterfactual welfare.

For maximum counterfactual welfare, we set $\beta = 0$ and make the obedience constraint for not entering hold as an equality. Intuitively, we increase welfare by having firms enter less often, so as to avoid the low-payoff from duopoly. Solving for γ, we obtain

$$\gamma = 1 - \frac{1}{1 + z} = \frac{z}{1 + z}.$$

So γ goes from 0 to $1/2$ as z goes from 0 to 1. Note that when $\beta = 0$, the obedience constraint for entering is unambiguously satisfied, since the left-hand side reduces to $1/2$, and the right hand side is always at least 1/2. The resulting payoff is

$$2(3+z)-4-2\frac{z}{1-z},$$

which coincides with the simulation.

We next consider the equilibrium to enter if and only if $c_i = 0$. The payoff from entering with a low cost is clearly positive. The payoff from entering with the high cost is just z, and the payoff from entering with a low cost is $2+z$, so this is an equilibrium if $z \in [-2, 0]$. The resulting ex ante sum of payoffs is

$$\frac{3+z}{2}+\frac{1+z}{2}=1+z.$$

This is the unique counterfactual prediction when $z \in (-1, 0)$, and it is the lower boundary of the counterfactual prediction when $z \in [-2, -1)$.

If $z \in [-3, -2]$, there is an equilibrium in which low-cost firms mix over whether they enter, which results in a payoff of zero. This attains the lower boundary of the counterfactual prediction for $z \in [-3, -2]$.

Next we construct the producer surplus maximizing BCE when $z \in [-3, -1]$. Using a correlation device as we did above for $z \in [0, 1]$, we can coordinate the low firms' behavior so that firms enter only if they have low cost, a firm enters with probability one if they are the only low-cost firm, and when both firms have low-cost, and exactly one firm enters when both firms have low cost. This is obviously an equilibrium: Entering is strictly dominated for high signals, and if a firm with low cost does not enter in equilibrium, then the other low-cost firm must be entering, so the payoff from deviating would be $1+z \leq 0$. The resulting aggregate payoff would be $3(3+z)/4$ (that is, 3/4 of the time exactly one firm enters, and it is a firm with low cost). This coincides with the upper boundary of the simulation.

Finally, we construct an equilibrium that attains the low payoff at $z = -1$. First, there is a correlation device as above when $\gamma = 1/2$. In addition, we assume that low-cost firms can observe the cost of the other firm. Consider the following strategies: A high-cost firm enters if and only if $s_i = 1$. A low cost firm enters with probability 1 if the other firm's cost is low or if the other firm's cost is high and $s_i = 1$. Otherwise, when the other firm's cost is high and $s_i = 0$, the low-cost firm does not enter. The high-cost firm gets zero surplus from entering. Relative to the equilibrium where firms enter if and only if the cost is low, producer surplus has dropped by 1/2, since 1/4 of the time it is a high cost firm entering as a monopolist rather than a low-cost firm. This equilibrium is knife edge: First, it depends on the fact that the loss from duopoly is the same as the high entry cost, so that the low-cost firm is indifferent to entering as a duopolist, and the high-cost firm is indifferent to entering as a monopolist. Second, if z is a little bigger than 1, low-cost firms would strictly prefer to enter when the high-cost firm enters, and if z is a little smaller than 1, the high-cost firm would be unwilling to enter.

Informationally-Robust Rankings in the Entry Game In this appendix, we conduct version of the joint counterfactual prediction analysis described in Section 7. In this case, we ask whether higher z are necessarily associated with higher payoffs for the firms. We conducted six versions of this counterfactual, which are depicted in Figure 6.

We computed joint predictions for counterfactual producer surplus for two different counterfactual games: $z = -0.6$ and $z = -0.4$. Three versions of this computation under different informational assumptions are depicted in Figure 6.

First, we computed a counterfactual prediction when we restrict attention to information structures that can rationalize the data used in Section 4.2, namely that when $z = 0$, firms enter if and only if their cost is low. Thus, in this example, we are actually computing joint predictions for three games, where $z \in \{-0.6, -0.4, 0\}$, and we impose a data restriction on the $z = 0$ game and plot the set of pairs of counterfactual producer surplus for the games with $z = -0.6$ and $z = -0.4$. In fact, for this case, the set of joint counterfactual predictions can immediately be read from Figure 2: For $z \in (-1, 0)$, there is a unique counterfactual prediction for aggregate payoffs under fixed information, and this prediction is increasing in z. Indeed, we see in Figure 6 that the joint counterfactual prediction when we have the data restriction is the single blue point. This point is above the 45 degree, meaning that the firms are unambiguously better off in the aggregate when $z = -0.4$ than they are when $z = -0.6$.

Second, for these same parameters, we computed a joint prediction when we impose that the same information structure is used for both values of z, but we allow all private-cost information structures (depicted in red). In this case, the joint prediction spans both sides of the 45 degree line, so that $z = -0.4$ does not dominate $z = -0.6$ with fixed information, when we do not have a restriction from the data. A fortiori, $z = -0.4$ does not dominate $z = -0.6$ under unrestricted information, even when we restrict to private-cost information structures.

Third, we computed the joint prediction when we allow all information structures. This most permissive joint prediction for producer surplus is in green. Again, it is clear that neither game dominates the other.

This example illustrates the potential benefit of combining methodologies: When we use only joint predictions for informationally-robust rankings, without a data-based restriction, it is not possible to rank $z = -0.6$ and $z = -0.4$. But when we use data to further refine the joint counterfactual prediction, we do obtain an unambiguous ranking.

8.3 Two-Player Zero-Sum Game

We now consider a setting with two players, binary actions, and binary states. The observed game is the following:

$\theta = 0$			$\theta = 1$		
a_1/a_2	0	1	a_1/a_2	0	1
0	$(2, -2)$	$(-1, 1)$	0	$(0, 0)$	$(-1, 1)$
1	$(-1, 1)$	$(0, 0)$	1	$(-1, 1)$	$(2, -2)$

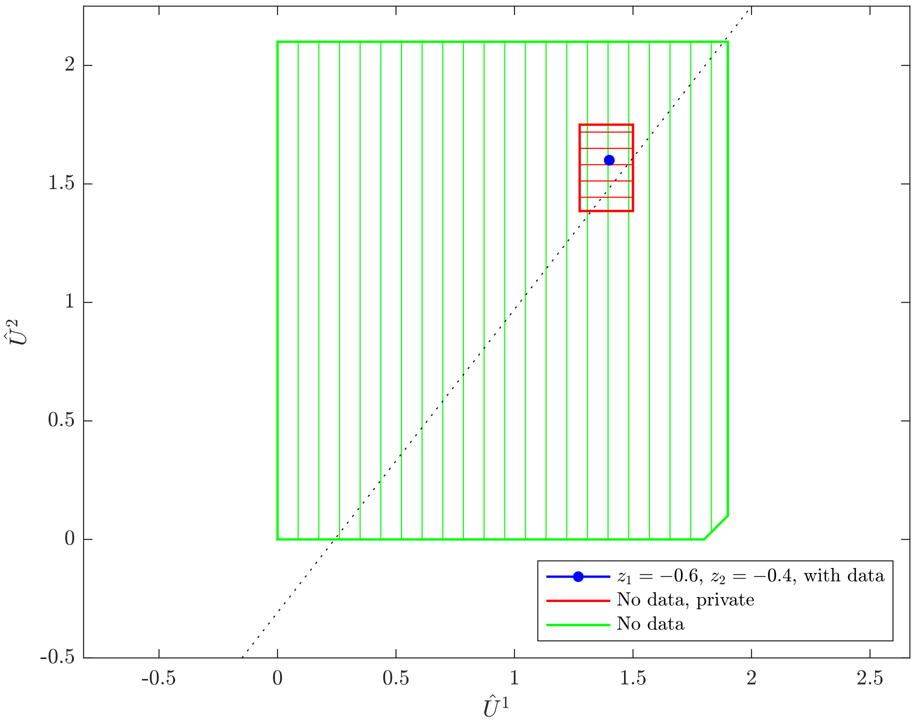


Figure 6: Joint counterfactuals in the entry game.

In each state, the game has the form of an asymmetric matching pennies. Both states are equally likely, so that in expectation the game is symmetric. Thus, if the players have no information about the state, there is a unique equilibrium in which they both randomize with equal probabilities, and both players' payoffs are zero. If they have full information about the state, then there is again a unique (and symmetric) equilibrium in which they play $a = 0$ with probability 1/4 in state $\theta = 0$, and they play $a = 0$ with probability 3/4 in state $\theta = 1$. In both states, player 1's payoff is $-1/4$.

We assume that we have observed ϕ exactly, and $\phi(a, \theta) = 1/8$ for all (a, θ). This is the joint distribution of states and actions that arises under no information. In the counterfactual, we multiply all of the payoffs by a factor $2 - z$ in state 0 and by z in state 1, for some $z \in [0, 2]$. This is equivalent to varying the relative likelihoods of the two states. The observed game corresponds to $z = 1$. The counterfactual outcome of interest is player 1's payoff.

We numerically computed maximum and minimum payoffs for player 1 for a fine grid of z values. The range of counterfactual outcomes under variable and fixed information are depicted in Figure 7 as a function of z. When information is variable, then again, the only thing we learn from the data is that both states are equally likely. The gray lines represent upper and lower bounds on welfare. The range of possible outcomes is largest at $z = 1$, when the counterfactual game is a copy of the observed game. In this case, any payoff in $[-1/2, 1/2]$ can be attained with some information structure. The highest payoff of 1/2 can be achieved by letting player 1 observe the state and player 2 receiving no information. Under that information, there is an equilibrium where $\widehat{a}_1 = \theta$ and player 2 mixes with equal probabilities. Similarly, the payoff of $-1/2$ can be achieved by giving no information to player 1 and full information to player 2. In fact, it is a result of Peski (2008) that these are the information structures that achieve extreme welfare outcomes in any two-player zero-sum game, and it is not particular to our example.[23]

[23] Here is a sketch of the proof. Player 1's payoff in $(\mathcal{G}, \mathcal{I})$ is at least their maxmin payoff, where the max and min are taken over player 1 and player 2's strategies, respectively. Player 2 has the option to use a

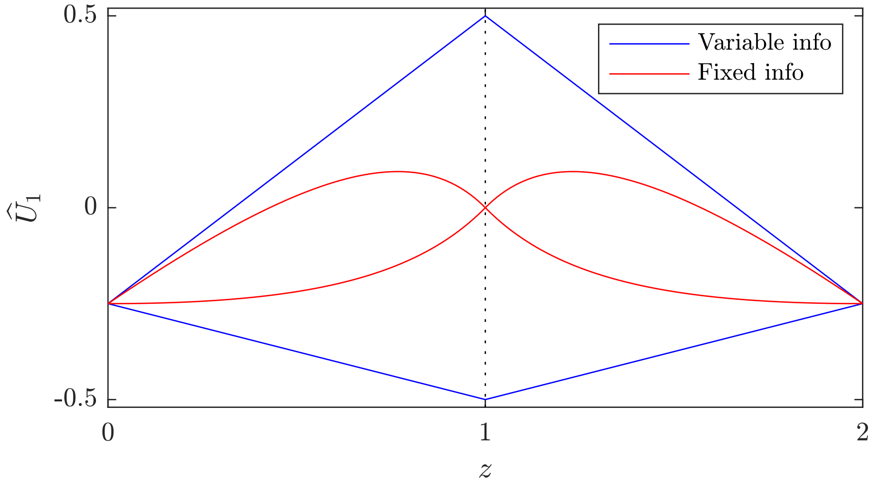


Figure 7: Counterfactual payoffs for player 1 in the zero-sum game.

Note that when $z = 0$ or $z = 1$, then payoffs are zero in one state, so that it is effectively a game with a single state, and thus the value of the game is uniquely pinned down independent of the information.

When we fix information, the range of counterfactual outcomes is tighter. Indeed, when $z = 1$, there is a unique counterfactual prediction when the counterfactual game coincides with the observed game. Once again, this is a general insight that is not particular to our example. In any two-player zero-sum game, if there is an information structure $\mathcal{I}$ and equilibrium σ that rationalizes the observed actions and in which player 1's payoff is u_1, then it must be that the zero-sum game $(\mathcal{G}, \mathcal{I})$ has a value which is u_1, and hence all equilibria have the same payoffs. This observation completes an analogue of Proposition 3 for zero-sum games:

strategy that does not depend on their private information t_2, so player 1's maxmin payoff would increase if we restricted player 2 to use only those constant strategies. This is what happens if player 2 has no information. Next, if we look at information structures where only player 1 gets information, then it must be that player 1's payoff is maximized by having as much information is possible. For, any strategy under partial information can be replicated under full information simply by "simulating" the noisy signal, so the effective strategy space is largest under full information. Finally, in the extreme case of full information/no information, the game is finite so the minimax theorem holds, and the maxmin payoff is player 1's equilibrium payoff.

Proposition 4 (Two-Player Zero-Sum Counterfactuals)**.**
Consider a two-player zero-sum game in which players' observed payoffs are $(u_1, -u_1)$*. If the counterfactual and observed games are the same, then under fixed information, there is point identification of the players counterfactual payoffs, which must be* $(u_1, -u_1)$*. Under unrestricted information, then a tight upper bound on player 1's payoff is given by what is attained when player 1 has full information and player 2 has no information, and a tight lower bound is what is attained when player 1 has no information and player 2 has full information.*

Thus, it is a general phenomenon that there are point predictions for local counterfactuals in two-player zero-sum games under fixed information, although there is generally a fat set of counterfactual predictions under unrestricted information.

Returning now to the particular example, as z moves away from 1, the range of counterfactual payoffs expands, before contracting again as we approach the complete information extremes. Thus, the predictive power of fixed information is large when the counterfactual is close to the observed game, and it degrades as the counterfactual environment diverges from that which generated the data.

The broad economic conclusion is that player 1 prefers moderate z, while player 2 prefers extreme values. Specifically, when information is fixed and $|z - 1| > 0.58$, then we can unambiguously say that player 1 is worse off and player 2 is better off in the counterfactual than in the observed game. When $|z - 1| \leq 0.58$, then the change in welfare is ambiguous: player 1 may be better off or worse off, depending on the true information structure. A similar statement applies when information is variable, but the conditions for player 1 to be better off are more stringent, and we can unambiguously sign the change in welfare only when $|z - 1| > 2/3$.

8.4 First-Price Auction

Our final example is a private-values first-price auction (cf. Section 5). This setting is similar to the one initially studied by Syrgkanis et al. (2021), except that we consider counterfactuals with fixed information, whereas they allow unrestricted information that there are two bidders with values in $V = \{0, 1/9, \ldots, 8/9, 1\}$. We also restrict bids to be in the value grid, and we also assume that bidders do not bid more than their values. There is no reserve price in the auction. Bidders learn at least their own value, but may learn more. We assume that the values are iid uniform, and the econometrician observes either the BCE that minimizes the auction's revenue or the BCE that maximizes revenue (both of which are computed numerically). The counterfactual of interest is revenue as we vary the reserve price. In particular, does there exist a reserve price under which revenue unambiguously increases, relative to the observed game without a reserve price?

Let us first consider the case where the observed outcome was the revenue minimizing BCE. Figure 8 shows how the counterfactual prediction for revenue varies with the reserve price. In particular, the solid red curves represent maximum and minimum counterfactual revenue. There are two features to notice: First, even if the reserve price stays at zero, there is a fat set of counterfactual revenue levels. This indicates that there exist information structures that could induce the revenue minimizing BCE for which there are multiple equilibria, and that revenue varies across these equilibria. So, even if the reserve price does not change, revenue could in principle increase if the bidders coordinated on a different equilibrium. This multiplicity persists at higher reserve prices. However, for moderate reserve prices, the lower bound on revenue increases above the observed level. This lower bound is maximized at 5/9. At this reserve price, we can unambiguously say that regardless of the information and equilibrium, revenue would necessarily be higher than in the observed outcome. Note that since the lowest value is zero, it is necessarily the case that minimum revenue increases when the reserve price changes from 0 to 1/9, although it is not obvious that revenue should continue to increase in the reserve price beyond this point.

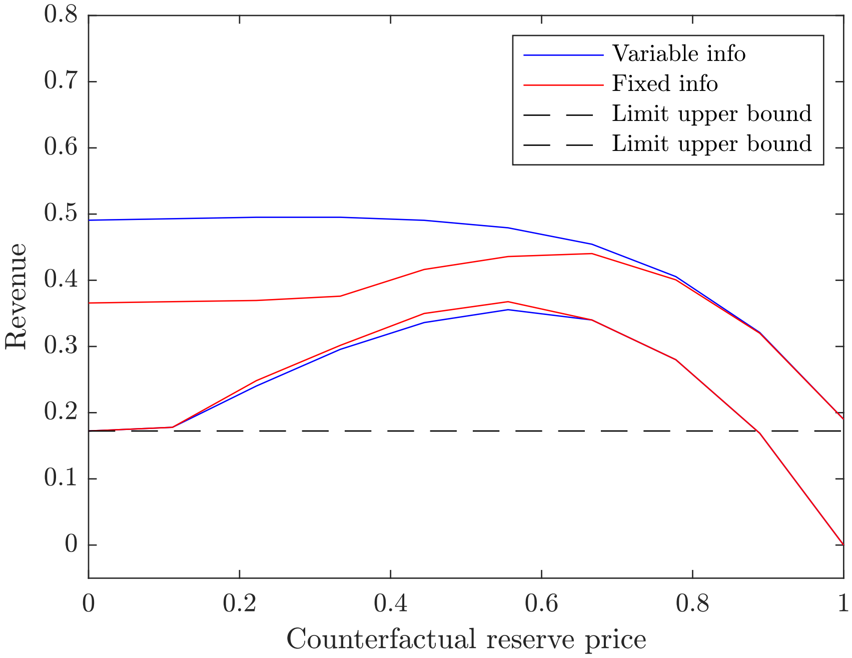


Figure 8: Counterfactual when observed outcome is the revenue minimizing BCE.

Figure 8 also shows how the counterfactual prediction if we allowed information vary, but held fixed the value distribution. For the lower bound, the predictions are not substantively different, although the upper bound on revenue is considerably more permissive. This is not surprising: The simulated data came from the revenue-minimizing information structure, so the fact that the lower red and blue curves nearly coincide is a reflection of the fact that the revenue-minimizing information does not vary significantly with the reserve price.

We next consider the case where the observed outcome is the revenue maximizing BCE. The corresponding counterfactual prediction is depicted in Figure 9. In this case, adding a reserve price cannot lead to a significant increase in revenue, and will necessarily cause revenue to decrease if it the reserve price is sufficiently high. Again, this prediction is substantively the same as what we would obtain with unrestricted information, although in this case it is the lower bound on revenue that is more permissive with unrestricted information. In fact, we can give an analytical justification for both the fact that maximum revenue is (nearly) decreasing in the reserve price, and also the fact that the fixed- and unrestricted-information bounds coincide. As discussed in Bergemann, Brooks, and Morris (2017, Section 5.4), under the hypothesis that bidders do not bid more than their values, there is an elementary lower bound on bidder surplus, which is the maximum payoff a bidder could obtain if others were bidding their values. With two bidders whose values are exactly uniformly distributed on $[0,1]$, and when the reserve price is r, the lower bound for a bidder with value $v \geq r$ is the maximum of

$$\max_{b \in [r,v]} (v-b)\, b = \begin{cases} \frac{v^2}{4} & \text{if } v \geq 2r; \\ (v-r)\, r & \text{if } r \leq v < 2r. \end{cases}$$

(If $v < r$, the lower bound on bidder surplus is zero.) The lower bound on ex ante bidder surplus when $r \leq 1/2$ is therefore

$$2\left[\int_{v=r}^{2r} (v-r)\, r dv + \int_{v=2r}^{1} \frac{v^2}{4} dv\right] = \frac{1}{6} - \frac{r^3}{3},$$

and when $1/2 \leq r \leq 1$, the lower bound is

$$2\int_{v=r}^{1} (v-r)\, r dv = \left(v^2 - 2rv\right) r\Big]_{v=r}^{1} = r - 2r^2 + r^3.$$

At the same time, total surplus when the reserve price is r is at most the expected highest value times an indicator for the highest value being above r, which is

$$\int_{v=r}^{1} vd\left(v^2\right) = \frac{2}{3}\left(1-r^3\right).$$

Thus, an upper bound on revenue with a reserve price r is

$$\overline{R}\,(r) = \frac{2}{3}\left(1-r^3\right) + \begin{cases} \frac{r^3}{3} - \frac{1}{6} & \text{if } r < \frac{1}{2}; \\ 2r^2 - r - r^3 & \text{if } \frac{1}{2} \leq r \leq 1. \end{cases}$$

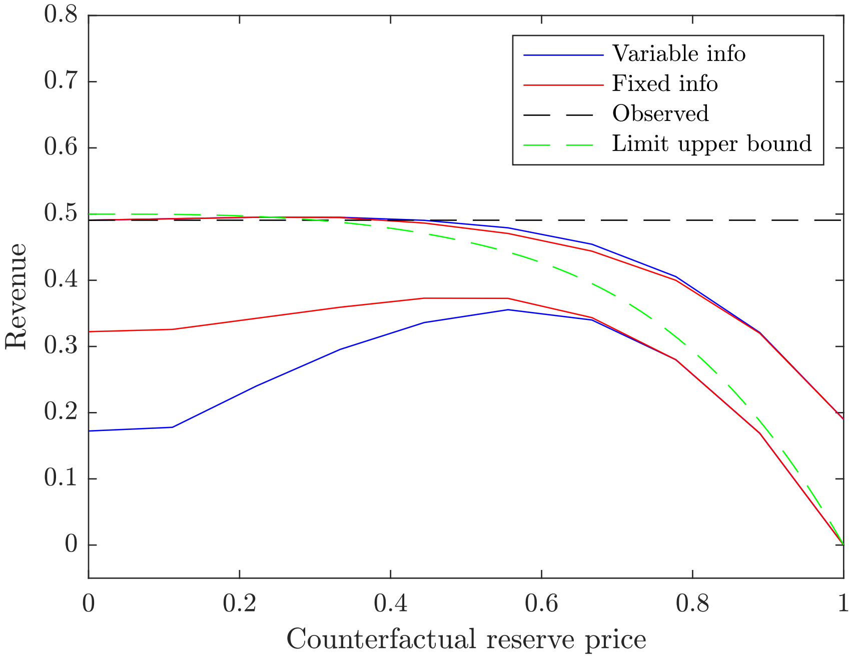


Figure 9: Counterfactual when observed outcome is the revenue-maximizing BCE.

We have plotted $\overline{R}$ in green in Figure 9. It is straightforward to verify that this function is decreasing. Moreover, Bergemann, Brooks, and Morris (2017) show that the bound is tight when $r = 0$, meaning that there exists an information structure and equilibrium in which revenue is $\overline{R}(0)$, so that in the limit when the value and bid grids fill in all of $[0, 1]$, maximum revenue must be decreasing in the reserve price. We conjecture that this construction can be generalized to $r > 0$, so that in fact $\overline{R}(r)$ is maximum revenue across all BCE.

As a final note, this counterfactual exercise extracts as much from the data as possible about players' information, as it pertains to this particular counterfactual prediction. We may contrast this approach with one suggested by us in our analysis of BCE of interdependent value first-price auctions (Bergemann, Brooks, and Morris, 2017). In that paper, we identified a tight lower bound on the winning bid distribution across all BCE consistent with a given ex post value distribution. We suggested using that bound to partially identify the value distributions that can rationalize observed winning bids. This partially identified set could then be used to generate counterfactual predictions. Such an exercise would allow information to vary between the observed and counterfactual auctions. In contrast, the methodology in the present paper holds information fixed between observation and counterfactual. Our focus is also less on the identification of values than on the identification of information, although we could have also treated the value distribution as a latent variable to be identified from the BCE, in which case we would have been using the entire observed bid distribution to implicitly restrict the value distribution, rather than just the distribution of the winning bid.

8.5 Innocuous Assumptions

Our model imposes a great deal of structure on the environment. In particular, we have assumed that information is described by a single information structure, utilities are known, the prior over the state is held fixed, and there is a single equilibrium that is played in the observed game and a single equilibrium in the counterfactual. At first glance, this structure seems restrictive for empirical applications in which the data is generated by many different instances from the observed game, and where conditions may vary from one instance to another. But, as we will now explain, these assumptions are without loss of generality and could be relaxed at the expense of a richer model.

1. All players receive signals from the same information structure. In practice, players with different characteristics, in different locations, or different points in time may receive qualitatively different forms of information. We may, however, consider these to be special cases of global description of players' information, where the heterogeneity in information is encoded as an extra dimension of signal. For example, suppose that for each $k = 1, \ldots, K$, a fraction $\beta_k \in [0,1]$ of the data is generated when the players have common knowledge that the information structure is $\mathcal{I}^k = \left\{S_1^k, \ldots, S_n^k, \pi^k\right\}$. We could equivalently represent this economy with a new information structure in which $S_i = \sqcup_{k=1}^{K} \{k\} \times S_i^k$, i.e., each player's set of signals is a disjoint union of the k information

structures, and

$$\pi\left(X,\theta\right)=\begin{cases}\beta_k\pi^k\left(Y,\theta\right) & \text{if } X=\{k\}\times Y \text{ for some } k;\\ 0 & \text{otherwise.}\end{cases}$$

In words, with probability one, all players get signals in the same S^k, and each k has probability β_k. Our counterfactual prediction implicitly allows for information structures of this form.

2. The utility functions $u_i\left(a,\theta\right)$ are known to the analyst. Uncertainty about preferences can be incorporated by expanding the state space. For example, suppose we start with a state space Θ, a moment restriction $M=\{\phi\left(a,\theta\right)\}$, and two possible utility functions u^1 and u^2. Then we can expand the state space to $\widetilde{\Theta}=\{1,2\}\times\Theta$, utility function $u\left(a,\left(k,\theta\right)\right)=u^k\left(a,\theta\right)$, and the moment restriction is

$$M=\left\{\widetilde{\phi}\in\Delta\left(A\times\widetilde{\Theta}\right)\middle|\sum_{k=1,2}\widetilde{\phi}\left(a,\left(k,\theta\right)\right)=\phi\left(a,\theta\right)\right\}.$$

Thus, the prevalence of u^1 and u^2 in the population is a free variable, and is partially identified from the data.

3. The distribution over states μ is held fixed in the counterfactual. In fact, we can allow a different distribution $\widehat{\mu}$ in the counterfactual, as long as it is absolutely continuous with respect to μ, meaning that it can be written as $\widehat{\mu}\left(\theta\right)=\eta\left(\theta\right)\mu\left(\theta\right)$ for some $\eta:\Theta\to\mathbb{R}_+$, and the conditional distribution of signals remains the same, meaning that the joint distribution of signals and states in the counterfactual is $\hat{\pi}\left(ds,\theta\right)=\eta\left(\theta\right)\pi\left(ds,\theta\right)$. In particular, when we are only interested in varying the prior and the absolute continuity hypothesis is satisfied, then we can set the counterfactual utility

to $\widehat{u}_i(a,\theta) = \eta(\theta) u_i(a,\theta)$, in which case equilibrium utility is simply

$$\begin{aligned}\sum_{\theta\in\Theta}\int_{s\in S}\sum_{a\in A}\widehat{u}_i(a,\theta)\,\sigma(a|s)\,\pi(ds,\theta) &= \sum_{\theta\in\Theta}\int_{s\in S}\sum_{a\in A}\eta(\theta)\,u_i(a,\theta)\,\sigma(a|s)\,\pi(ds,\theta)\\ &= \sum_{\theta\in\Theta}\int_{s\in S}\sum_{a\in A}u_i(a,\theta)\,\sigma(a|s)\,\hat{\pi}(ds,\theta),\end{aligned}$$

and the represented payoffs are equivalent to those that would obtain with the different prior. This is merely a reflection of the well-known indeterminacy of probabilities versus utilities in the subjective expected utility model, when utilities are state dependent (Savage, 1954; Anscombe and Aumann, 1963). Indeed, this transformation was being used in the single-player analysis of Section 4.1, which can be reinterpreted as variations of the prior.

4. All players play the same equilibria of the observed and counterfactual games. This is also without loss of generality. Suppose that the information structure is $\mathcal{I}$, and a share β_k of the data is generated from players who play strategies σ^k for $k = 1, \ldots, K$. The same outcome can be induced with a single information structure $\widetilde{\mathcal{I}}$, in which $\widetilde{S}_i = \{1, \ldots, K\} \times S_i$, $\widetilde{\pi}(\{k\} \times X, \theta) = \beta_k \pi(X, \theta)$, and strategies are $\tilde{\sigma}_i(a|(k,t)) = \sigma_i^k(a|s)$. In effect, the first coordinate of the new signal $\widetilde{s}_i$ is a public randomization device which is equal to k with probability β_k. Strategies on the larger space say to play σ^k when $X = k$.