Counterfactual Analysis in Bargaining with Externalities: A Matching-Theoretic Foundation*

Marzena Rostek† and Nathan Yoder‡

March 20, 2026

Abstract

We show how matching-theoretic stability can be used to perform counterfactual analysis in environments with externalities. Our approach sidesteps the well-known existence problems faced by stability in these environments by endowing agents with correct beliefs about the choices of others. This facilitates a procedure that mirrors the one commonly used with the Nash-in-Nash solution concept, with beliefs playing the same role as bargaining weights: Like profiles of bargaining weights in Nash-in-Nash, each profile of beliefs predicts a unique outcome, but different profiles predict different outcomes.

Keywords: Externalities, matching with contracts, stability, Nash-in-Nash

1 Introduction

A great deal of recent empirical work explores settings where agents form agreements with one another that have externalities on other agents. In industrial organization, researchers have studied agreements between cable television producers and distributors (Crawford and Yurukoglu (2012)) or insurers and healthcare

* Some parts of this paper were previously included in "Matching with Strategic Consistency". The authors are grateful to numerous colleagues for their helpful comments and suggestions.

† University of Wisconsin-Madison, Department of Economics; E-mail: mrostek@ssc.wisc.edu.

‡ University of Georgia, Terry College of Business, John Munro Godfrey, Sr. Department of Economics; E-mail: nathan.yoder@uga.edu.

providers (Ho and Lee (2017)); in international economics, trade agreements between countries (Bagwell et al. (2021)). A central focus in this literature is *counterfactual analysis* describing how outcomes would change in response to some exogenous change in the setting. For instance, Ho and Lee (2017) estimate the effects of the removal of an insurer from the healthcare market, while Bagwell et al. (2021) estimate the effects of a change in GATT/WTO rules.

In this paper, we show by example how our results from Rostek and Yoder (2025) can be applied to perform counterfactual analysis in settings like these. Because the approach that we take is based on the canonical concept of *stability* from matching theory, its predictions are robust to deviations to *arbitrary* sets of agreements. For instance, firms can consider swappping an agreement with one supplier for an agreement with another (or multiple other suppliers), or simultaneously consider agreements with multiple suppliers for the purchase of complementary inputs. Assumptions about bargaining protocols, delegated agents, or exogenously fixed outside options are unnecessary.

Our approach to this exercise mirrors the approach often taken with the *Nash-in-Nash* (Horn and Wolinsky (1988)) solution concept — a Nash equilibrium in Nash bargains. This solution concept has found extensive use in both theoretical and empirical work. Nash-in-Nash offers a key advantage that has helped make it popular: It avoids the nonexistence issues faced by stability in applications with externalities by considering a smaller class of deviations. As we show in Rostek and Yoder (2025), endogenizing agents' choices as best responses to beliefs about the choices of others allows matching-theoretic stability to overcome the challenges for existence of a stable outcome in these environments, while still allowing agents to renegotiate agreements in arbitrary ways.

For instance, if an upstream supplier has economies of scale (e.g., a hospital (Dranove, 1998)), it may increase its profit by lowering the prices it has agreed to with *multiple* downstream firms (e.g., insurers), even when lowering a *single* price would not benefit it. Alternatively, if a retailer wishes to make a "captaincy" agreement for the majority of its shelf space with a single supplier (as in, e.g., Clark et al. (2024)), its relevant choice is *between* suppliers, not just different quantities from the same supplier. Moreover, such agreements with different suppliers are mutually exclusive: some retailer-supplier pairs must be excluded from the outcome. These incentives to deviate from agreements with multiple counterparties — and to form agreements with some counterparties, but not others — are captured by stability, but not by Nash-in-Nash.[1]

Endogenizing agents' choice functions as responses to correct beliefs — ensuring that agents' choices are *optimal* given *beliefs* about others' choices that are *correct* and *consistent* across sets of alternatives —creates a degree of freedom that is similar to the one exploited by Nash-in-Nash counterfactual analysis. Observe that many outcomes are predicted by Nash-in-Nash for some bargaining parameters, but any given vector of bargaining parameters produces a unique Nash-in-Nash outcome. Thus, observed outcomes can be used to recover the object (a vector of bargaining parameters) that selects them, and that object can be used to select a new outcome after an exogenous change. Likewise, we show in Rostek and Yoder (2025) that each profile of correct, consistent beliefs pins down a *unique* stable outcome. Because of the fixed point relationship between choices and beliefs, there may be several such profiles. This suggests exploiting this degree of freedom in the same way, by recovering a profile of beliefs from data and then using it to pin down a new outcome after an exogenous change in the model.

This is what we do in this paper. As Rostek and Yoder (2025) shows, this analogy between beliefs and bargaining weights can be formalized: Correct, consistent beliefs can be constructed by maximizing an asymmetric Nash product, subject to constraints that capture the agents' (endogenous) outside options. We can thus think of these beliefs as a microfoundation for Nash bargaining weights, or vice versa. Since these beliefs each pin down a unique stable outcome, we show that we can use this maximization problem in exactly the same way that the Nash-in-Nash maximization problems are used in practice for counterfactual analysis. Thus, we can pin down the impact on beliefs and outcomes of, e.g., changes to the set of agreements (e.g., disallowing an agreement through regulation, or requiring products to be sold through an intermediary), changes to the set of agents involved in an agreement (e.g., requiring the approval of nearby hospitals to build a new one), or changes to the agents' payoffs (e.g., through common ownership of two firms, the imposition of a tax, or the introduction of other types of externalities).

[1] Note, however, that authors such as Ho and Lee (2019) have modified the baseline Nash-in-Nash solution concept to allow for exclusion.

2 Environment

We demonstrate our approach to stability and counterfactual analysis in a model of contracting between upstream firms (i.e., suppliers of inputs) $i \in I$ and downstream firms (i.e., producers of final goods) $h \in H$ à la Collard-Wexler et al. (2019). A *contract* is an agreement between an upstream firm i and a downstream firm f that specifies a price (or more generally, a vector of prices) $p_{if} \in \mathcal{P} \subset \mathbb{R}_+$. These prices may be lump sum payments, as in Collard-Wexler et al. (2019), or prices per unit of input sold (e.g., price per subscriber for each channel in Crawford and Yurukoglu (2012), or price per stent in Grennan (2013)). The set of all such contracts is thus $X := \{(i, f, p_{if}) : i \in I, f \in F, p_{if} \in \mathcal{P}\}$. For each upstream firm $i \in I$, we let $X_i := \{(i, f, p_{if}) : i \in I, p_{if} \in \mathcal{P}\}$ denote the set of contracts involving i, and $X_{-i} := X \backslash X_i$ denote the set of contracts that do not; likewise, for each downstream firm $f \in F$, we let $X_f := \{(i, f, p_{if}) : i \in I, \tau_{if} \in \mathcal{P}\}$ denote the set of contracts involving f, and $X_{-f} := X \backslash X_f$ denote the set of contracts that do not. Similarly, for sets of contracts $Y \subseteq X$, we write $Y_j := Y \cap X_j$ and $Y_{-j} := Y \cap X_{-j}$ for each $j \in I \cup F$.

Given a set of contracts $Y \subseteq X$ — an *outcome* — downstream firms compete with one another in the market for final goods.[2] Their equilibrium pin down the payoff $u_i(Y) > 0$ of each upstream firm i, and the payoff $u_f(Y) > 0$ of each downstream firm f, from that outcome. These preferences may exhibit externalities: in general, $u_j(Y)$ depends on Y_{-j}, not just Y_j. For instance, in Crawford and Yurukoglu (2012), the number of subscribers to a television channel i through a distributor f, and the prices that they pay, depend on the equilibrium bundles and pricing schemes offered by all distributors in that market. Since that equilibrium depends on the *entire* set of contracts between TV producers and distributors, so do producer and distributor profits.

Each manufacturer-hospital pair can only sign at most one contract; we capture this by setting $u_j(Y \cup \{(i, f, p_{if}), (i, f, p'_{if})\}) < u_j(Y \cup \{(i, f, p_{if})\})$ for each $i \in I$, $f \in F$, and $j \in \{i, f\}$, and each pair of prices $p_{if}, p'_{if} \in \mathcal{P}$. When an outcome contains at most one contract between each pair of firms, a firm j's payoff is just its profit $u_j(\{(i, f, p_{if})\}_{(i,f) \in A}) = \pi_j(\mathbf{p}_A; A)$ given the vector $\mathbf{p}_A$ of prices agreed to by the upstream-downstream pairs $A \subseteq I \times F$ that contract with each other. We assume

[2] As in Collard-Wexler et al. (2019), we abstract away from the form of that competition to focus on bargaining between upstream and downstream firms.

that each π_j is continuous.

2.1 Stability

We use stability, the canonical solution concept in matching theory, extended to allow for externalities. Whether an outcome is stable or not depends on preferences indirectly through the agents' *choice functions* $C_j : 2^{X_j} \times 2^{X_{-j}} \to 2^{X_j}$. Firm j's choice function C_j takes two arguments: the set of contracts that are available (i.e., under discussion in a negotiation) that involve firm j, and the set of contracts that are available but do not involve firm j. $C_j(Y_j|Y_{-j}) \subseteq Y_j$ is the set of contracts that firm j chooses when Y is the set of contracts under discussion.

Definition. Given choice functions $\{C_j\}_{j\in I\cup F}$, a set of contracts $Y \subseteq X$ is stable if it is

i. *Individually rational*: $Y_j = C_j(Y_j|Y_{-j})$ for all $j \in I \cup F$.

ii. *Unblocked*: There does not exist a nonempty $Z \subseteq (X\backslash Y)$ such that for all $j \in I \cup F$ with $Z_j \neq \varnothing$, $Z_j \subseteq C_j((Z \cup Y)_j|(Z \cup Y)_{-j})$.

In words, a set of contracts is stable if (a) when it is available, no one rejects any contracts from it (individual rationality), and (b) no group of agents can propose a new set of contracts, or *block*, that they are each willing to choose when made available (i.e., under negotiation) alongside the existing set of contracts.

2.2 Choices and Beliefs

Because pricing agreements have externalities on other manufacturers, stability could be difficult to apply with the standard, nonstrategic approach to choice, in which firms' choices $C_j(Y_j|Y_{-j})$ are simply given by their *favorite* subset of Y_j, without regard for whether some of those agreements might be rejected by others. This is because of well-understood existence issues that are not present with the Nash-in-Nash solution applied in papers such as Crawford and Yurukoglu (2012), Grennan (2013), or Ho and Lee (2017).

Rostek and Yoder (2025) introduces an approach that allows us to get around these issues and use stability to make predictions in environments with externalities. Its key innovation is to endogenously determine the agents' choice functions as part of a fixed point together with *correct* beliefs about the choices of others, such that beliefs are *consistent* across different sets of contracts that might be under discussion in a negotiation. Formally, an agent's beliefs are described by a function $\mu_i : 2^X \to 2^X$ that takes the set of available contracts as its argument, and returns the set of contracts that agent i believes will not be rejected by any of the other agents.

Definition. Given firms' payoff functions, the profile of choice functions and beliefs $\{C_j, \mu_j\}_{j \in I \cup F}$ is *strategically consistent* if for each $j \in I \cup F$,

i. μ_j is *correct* given $\{C_k\}_{k \neq j}$: For each $Y \subseteq X$,

$$\mu_j(Y) = \{(i, f, p_{if}) \in Y \mid (i, f, p_{if}) \in C_i(Y_i|Y_{-i}) \cap C_f(Y_f|Y_{-f})\}. \tag{1}$$

ii. C_j is *optimal* given μ_j: For each $Y \subseteq X$,

$$C_j(Y_j|Y_{-j}) \in \arg\max_S u_j(S \cup \mu_j(Y)_{-j}) \text{ s.t. } S \subseteq \mu_j(Y)_j. \tag{2}$$

iii. μ_j is *cross-set consistent* given $\{C_k\}_{k \in I \cup F}$: For each compact $Y, Z \subseteq X$, if $Y \supseteq Z \supseteq C_j(Y_j|Y_{-j})$ for all $j \in I \cup F$, then $\mu_j(Z) = \mu_j(Y)$.

The compactness criterion in (iii) is a technical condition ensuring that the implications of cross-set consistency are well-behaved even when, unlike in Rostek and Yoder (2025), sets of available contracts are not finite.

We focus on profiles with beliefs that satisfy the *Pareto optimality* refinement introduced in Rostek and Yoder (2025). This refinement ensures that agents do not believe that others will choose a Pareto-dominated set of contracts just because of coordination failure. Formally, suppose that Y is a *nonstrategically individually rational* set of contracts: $u_j(Y) \geqslant u_j(S \cup Y_{-j})$ for all $j \in I \cup F$ and $S \subseteq Y_j$. Agent j's beliefs μ_j satisfy Pareto optimality if for any such Y, there is no Z that (a) Pareto improves upon Y; (b) is also nonstrategically individually rational; but (c) j believes the other agents will still choose Y when Z has been proposed: $\mu_j(Y \cup Z) = Y$. If each agent's beliefs in a profile $\{C_j, \mu_j\}_{j \in I \cup F}$ satisfy Pareto optimality, we just say that the profile satisfies Pareto optimality.

Rostek and Yoder (2025) offers two results that are key to our framework for counterfactual analysis. First, these profiles of beliefs each pin down a unique stable outcome. Second, they can be generated by maximizing a social welfare function over the nonstrategically individually rational outcomes.[3] Proposition 1 extends these results to our setting where, unlike in Rostek and Yoder (2025), the set of all contracts X need not be finite, since prices may vary continuously (e.g., $\mathcal{P} = \mathbb{R}_+$).

Proposition 1. *Let $\phi : \mathbb{R}_+^{I \cup F} \to \mathbb{R}$ be a strictly increasing continuous function. There is a strategically consistent profile $\{C_j^\phi, \mu_j^\phi\}_{j \in I \cup F}$ with beliefs $\{\mu_j^\phi\}_{j \in I \cup F}$ that satisfy Pareto optimality such that for each j and each $Y \subseteq X$ such that $\{p_{if} | (i, f, p_{if}) \in Y\}$ is compact for each $(i, f) \in I \times F$,*

$$\mu_j^\phi(Y) \in \arg\max_{S \subseteq Y} \phi((u_j(S))_{j \in I \cup F}) \text{ s.t. } u_k(S) \geqslant u_k(S' \cup S_{-k}) \text{ for all } k \in I \cup F \text{ and } S' \subset S_k; \tag{SPP(Y)}$$

$$C_j^\phi(Y_j | Y_{-j}) \in \arg\max_S u_j(S \cup \mu_j^\phi(Y)_{-j}) \text{ s.t. } S \subseteq \mu_j^\phi(Y)_j.$$

Moreover, there is a unique outcome that is stable given the choice functions $\{C_j^\phi\}_{j \in I \cup F}$, and it solves (SPP(X)).

The role played by correct, consistent beliefs in our framework based on stability is analogous to that played by bargaining parameters in frameworks based on Nash bargaining: Each profile of beliefs or vector of bargaining parameters pins down a unique outcome, but different outcomes are predicted by different profiles of beliefs or bargaining parameters. This allows us to use them for our approach to counterfactual analysis analogously to the way that bargaining parameters are used in approaches based on Nash bargaining.

In fact, this analogy runs deeper: we can think of these beliefs as a microfoundation for Nash bargaining weights, or vice versa. Specifically, we can think of the welfare function ϕ used in Proposition 1 as formalizing the way that agents base their beliefs about other agents' choices on the payoffs the agents will receive from those choices. If this mapping from payoffs to beliefs is invariant under a rescaling of the firms' payoff functions, then it can be described by letting ϕ be an asymmetric Nash product $\phi(\mathbf{x}) = \prod_{j \in I \cup F} x_j^{\alpha_j}$ for some distribution $\alpha \in \Delta(I \cup F)$ (Lemma 3 in Rostek and Yoder (2025)). Thus, given the weights α, (SPP(X)) pins down the stable

[3] With a consistent tiebreaking rule, where necessary.

outcome $Y^*(\alpha)$ as

$$Y^*(\alpha) \in \arg\max_{S \subseteq X} \prod_{j \in I \cup F} u_j(S)^{\alpha_j} \text{ s.t. } u_k(S) \geqslant u_k(S' \cup S_{-k}) \text{ for all } k \in I \cup F \text{ and } S' \subset S_k. \quad (3)$$

It bears emphasizing that even though the planner's problem (3) constructs a stable outcome (and more generally, (SPP(Y)) constructs a strategically consistent profile of choice functions and beliefs) as the solution to a multilateral Nash bargaining problem, its interpretation is *not* that all firms bargain with one another á la Nash (1950) to *cooperatively* determine the outcome they will choose from each set of available contracts. Rather, at each set of available contracts, firms' choices are *individually* optimal given their beliefs about others' choices, and those beliefs are refined by some of the same axioms that characterize the Nash (1950) solution (Pareto optimality and scale invariance).[4] Given that profile of choice functions and beliefs, the stable outcome is then determined cooperatively by the absence of both individual and joint deviations. Hence, we can think of (3) as describing a "stability-in-Nash" approach: a stable outcome given optimal choices from beliefs that maximize a Nash product at each set of available contracts.[5]

3 Counterfactual Analysis

The procedure for Nash-in-Nash counterfactual analysis used in papers like Crawford and Yurukoglu (2012), Grennan (2013), or Ho and Lee (2017) can be described roughly as follows:

1. Use the data to recover (a) firms' preferences over agreements, and (b) the Nash bargaining weights that pin down the agreements actually observed. E.g., in Crawford and Yurukoglu (2012), estimate consumer demand as a function of the menu of television bundles and prices that they face, use equilibrium conditions from the competition in menus among the TV distributors to

[4] Though we do not explicitly invoke independence of irrelevant alternatives, this property of the Nash solution is precisely what ensures that the beliefs it generates satisfy cross-set consistency.

[5] By comparison, Nash-in-Nash is a "Nash equilibrium in Nash bargains": each pair of firms makes choices cooperatively from the agreements available to them, and their behavior is then determined noncooperatively across different pairs of firms.

recover firms' marginal costs, and finally estimate the bargaining weights for which the observed input prices are the Nash-in-Nash solution.

2. Use the recovered preferences and bargaining weights to make predictions in some counterfactual scenario. E.g., in Crawford and Yurukoglu (2012), modify the environment to account for a proposed regulation on the menus that TV distributors can offer, and then recompute the Nash-in-Nash prices.

Our approach follows the same steps, but instead of recovering Nash bargaining weights that pin down a bargaining solution, we recover *correct, consistent beliefs* that pin down a stable outcome. This may appear more challenging than the procedure used with the Nash-in-Nash concept. First, the observed outcome does not necessarily identify the full profile of beliefs and choice functions for which it is stable.[6] Second, unlike the set of vectors of bargaining weights, the set of strategically consistent profiles depends on — and thus is affected by exogenous changes to — the environment. Thus, we need a map from the set of profiles recovered in the first step to the set of profiles that are strategically consistent after an exogenous change occurs.

These difficulties can be overcome using the results from Rostek and Yoder (2025) described in Proposition 1. Recall from Proposition 1 that correct, consistent, Pareto optimal beliefs can be constructed by maximizing a social welfare function. Each such profile $\{\mu_j^\phi\}_{j\in I\cup F}$ can be identified with the welfare function ϕ used to construct it, which we can interpret as describing the way that agents base their beliefs about other agents' choices on the payoffs the agents will receive from those choices. Then, using the same ϕ, we can find *corresponding* beliefs and choices that are strategically consistent *after* an exogenous change in the environment. (Just as we assume that the same weights pin down the observed and counterfactual outcomes as part of the Nash-in-Nash procedure, we assume that the same ϕ describes the agents' beliefs before and after an exogenous shock.) Recovering the information about the strategically consistent profile necessary to make counterfactual predictions thus amounts to recovering the welfare function ϕ — or, when ϕ is pinned down by scale invariance, a vector of Nash weights α.

[6] That is, the same outcome can be stable for multiple strategically consistent profiles.

3.1 Stable Outcomes and Nash Weights

Both steps in the procedure outlined above require us to characterize the relationship between the stable outcome that we observe (or predict), and the welfare function that captures the way that agents' beliefs depend upon each other's preferences. We focus on welfare functions that are scale invariant, and can thus be represented as a Nash product. Thus, at a high level, this relationship is described by (3).

However, the structure of the model allows us to be more concrete. First note that since each upstream-downstream pair can only make one supply agreement, the outcomes that satisfy the individual rationality constraint in (3) cannot contain multiple agreements between the same pair of firms. Hence, we can write (3) as an optimization problem over (a) sets of firm pairs $A \subseteq I \times F$ that form agreements, and (b) price vectors $\mathbf{p}_A = (p_{if})_{(i,f)\in A}$ specified by those agreements, by replacing each firm's payoff function u_i over agreements with their profit function π_i over price vectors, as follows:

$$(A^*, \mathbf{p}^*_{A*}) \in \arg\max_{A,\mathbf{p}_A} \prod_{i \in I} \pi_i(\mathbf{p}_A; A)^{\alpha_i} \prod_{f \in F} \pi_f(\mathbf{p}_A; A)^{\alpha_f} \tag{4}$$

$$\text{s.t. } \pi_i(\mathbf{p}_A; A) \geqslant \pi_i(\mathbf{p}_{A\setminus B}; A\setminus B) \text{ for all } i \in I \text{ and } B \subseteq \{i\} \times F; \qquad (\text{NSIR}(i))$$

$$\pi_f(\mathbf{p}_A; A) \geqslant \pi_f(\mathbf{p}_{A\setminus B}; A\setminus B) \text{ for all } f \in F \text{ and } B \subseteq \{f\} \times I. \qquad (\text{NSIR}(f))$$

(4) describes the relationship between firm profits and beliefs (as characterized by the vector of bargaining weights α in the Nash product) and the supply relationships and negotiated prices that they pin down as part of a stable outcome. We can thus use it in the procedure for counterfactual analysis described in the outset of this section, in precisely the same way as the *set* of optimization problems that

characterize the Nash-in-Nash solution:[7]

$$p^*_{if} \in \arg\max_{p_{if}} \begin{array}{l} (\pi_f(p_{if}, \mathbf{p}^*_{I\times F\setminus\{(i,f)\}}; I \times F) - \pi_f(\mathbf{p}^*_{(I\times F)\setminus\{(i,f)\}}; (I \times F)\setminus\{(i,f)\}))^{\alpha_{if}} \\ \times(\pi_i(p_{if}, \mathbf{p}^*_{I\times F\setminus\{(i,f)\}}, I \times F) - \pi_i(\mathbf{p}^*_{(I\times F)\setminus\{(i,f)\}}; (I \times F)\setminus\{(i,f)\}))^{1-\alpha_{if}} \end{array} \tag{5}$$

for each $i \in I, f \in F$.

Using (4) in this way results in our procedure for counterfactual analysis based on stability (in the matching theory sense). We can first use (4) together with the data to recover the Nash bargaining weights that pin down observed agreements and prices. Then, we can make a change to the model and use the same bargaining weights to pin down agreements and prices in some counterfactual scenario.

3.2 Comparison to Other Approaches

We can offer a comparison of our approach with those based on the Nash-in-Nash bargaining solution, as well as with comparative statics based on the canonical nonstrategic approach to stability.

Comparison to Approaches Based on Nash-in-Nash

The contrast between (4) and (5) helps to illustrate the differences between our procedure and the usual one based on the Nash-in-Nash solution, and more generally, the differences between matching-theoretic stability and the Nash-in-Nash concept. First, stability endogenizes the set of firm pairs that form agreements. Consequently, it allows for endogenous exclusion: Unlike (5), (4) accommodates outcomes where some upstream-downstream pairs do not contract (and so $A^* \neq I \times F$). This could be because agreements between such pairs are not Pareto-improving, given the agreements signed by others. But it could also be because, due to the externalities that agreements have on other firms, a Pareto-improving set of contracts would not be individually rational. Thus, unlike the standard Nash-in-Nash procedure, our procedure permits analysis of the effect of a change in the environment (e.g., a

[7] See, e.g., Eq. 7 in Crawford and Yurukoglu (2012); Eq. 7 in Grennan (2013); Eqs. 3 and 4 in Ho and Lee (2017); or p. 174 in Collard-Wexler et al. (2019).

merger) on the network of agents that make agreements with each other, not just on the agreements formed among the agents in a fixed network.[8]

Example 1. Consider the following example adapted from Ho and Lee (2019). There is a single insurer (i.e., downstream firm) f, and two hospitals (i.e., upstream firms) j and k. Net of prices received from the insurer, hospital profits are constant: $\pi_i(\mathbf{p}_A; A) = 1 + p_{if}$ if $(i, f) \in A$, and 1 otherwise, for each $i \in \{j, k\}$. The insurer serves a single consumer who will be admitted to some hospital, and is willing to pay a higher premium for that hospital to be j:

$$\pi_f(\mathbf{p}_A; A) = \begin{cases} 1 - \sum_{i \in A} p_{if} + 10, & A \ni j; \\ 1 - p_{jf} + x, & A = \{k\}; \\ 1, & A + \varnothing. \end{cases}$$

Hence, nonstrategic individual rationality amounts to the requirement that either (a) the network A contains a single insurer-hospital pair, and the price paid to the hospital is less than the consumer's willingness to pay to be admitted to it (i.e., $A = \{(j, f)\}$ and $p_{jf} \in [0, 10]$, or $A = \{(k, f)\}$ and $p_{kf} \in [0, x]$); (b) $A = \{(j, f), (k, f)\}$ and $p_{kf} = 0$; or (c) $A = \varnothing$.

Suppose that bargaining parameters are symmetric across firms: $\alpha_f = \alpha_j = \alpha_k = \frac{1}{3}$. Clearly the Nash product cannot be maximized in case (c), and its value in case (b) is the same as its value in case (a) when $A = \{(j, f)\}$. When $A = \{(j, f)\}$, the Nash product is maximized at $p_{jf} = 5$; when $A = \{(k, f)\}$, the Nash product is maximized at $p_{kf} = \frac{x}{2}$; the former yields a higher value of the Nash product, so we have $(A^*, \mathbf{p}_{A^*}) = (\{(j, f)\}, 5)$. This outcome is stable for the strategically consistent profile with beliefs pinned down by equal bargaining parameters: Intuitively, even if $x > 5$, hospital k would not choose to block this outcome by accepting a price $p_{kf} < x - 5$, because it correctly believes that hospital j would just undercut it further: j's bargaining ability is just as strong as k's. This differs from the price ($p_{jf} = \min\{5, 10 - x\}$) predicted by Nash-in-Nash with threat of replacement for the set of contracting firms $A = \{(j, f)\}$ in Ho and Lee (2019), which assumes that

[8] The *Nash-in-Nash with Threat of Replacement (NNTR)* concept introduced by Ho and Lee (2019) allows for outcomes in which not every pair of firms contracts. In their approach, they use the observed outcome to pin down prices after an exogenous change in the set of firm pairs that contract with each other. Our approach instead allows both the set of firms that contract *and* the prices they agree to to be pinned down in a counterfactual environment.

hospital k would accept *any* nonnegative price, without considering the possibility of subsequent negotiations between f and j.

Alternatively, suppose that k's bargaining power is much higher: $\alpha_k = \frac{1}{2}$, but $\alpha_f = \alpha_j = \frac{1}{4}$. When $A = \{(j, f)\}$, the Nash product is still maximized at $p_{jf} = 5$, yielding a value of 36. When $A = \{(k, f)\}$, the Nash product is maximized at $p_{kf} = \frac{2x+1}{3}$, yielding a value of $4\left(\frac{2+x}{3}\right)^3$. If x is high enough — e.g., if $x > 5$ — the latter yields a higher value of the Nash product, so we have $(A^*, \mathbf{p}_{A^*}) = (\{(k, f)\}, \frac{2x+1}{3})$. That is, these bargaining weights pin down an outcome with inefficient exclusion, because Proposition 1 only guarantees that the outcome is efficient *subject to the constraint of individual rationality*: Even though a network of $\{(j, f), (k, f)\}$ would increase total surplus, the transfers that would turn it into a Pareto improvement on $(\{(k, f)\}, \frac{2x+1}{3})$ are not individually rational. Intuitively, unlike in the previous case, hospital j (correctly) does not believe that it could successfully undercut hospital k, because its bargaining ability is lower.

Second, with stability, agreements are not negotiated in isolation from each other. This is important whenever a firm can gain from modifying multiple agreements at once, but not from modifying those agreements individually. For instance, if an upstream firm has economies of scale in production, it may be willing to lower prices on a downstream firm — but only if it simultaneously agrees to lower the prices charged to other downstream firms. Alternatively, if a downstream firm views the products of two upstream firms as substitutable, it might profit by accepting a higher price from one and a lower price from the other. The existence of gains from simultaneous changes in multiple agreements impacts the stable outcome pinned down by (4), whereas in (5), gains from negotiation with different counterparties are considered separately.

Example 2. A downstream firm i uses two inputs, z_1 and z_2, which it purchases from two suppliers, $f \in \{1, 2\}$, with which it negotiates prices. For simplicity, we let the downstream firm's production function be quadratic: $f(\mathbf{z}) = \mathbf{a} \cdot \mathbf{z} - \frac{1}{2}\mathbf{z}'B\mathbf{z}$, where B is positive definite. Inputs are substitutes in production: B has nonnegative off-diagonal entries.[9] The suppliers produce at constant marginal cost c_f.

For $\mathbf{p}$ not too large, the firm's demand when it contracts with both suppliers is given by $D(\mathbf{p}) = B^{-1}(\mathbf{a} - \mathbf{p})$, and when it contracts only with firm f is given by

[9] That is, it is an M-matrix.

$\hat{D}_f(p_f) = \frac{1}{B_{ff}}(a_f - p_f)$. Thus, the firms' profit functions can be written (normalizing the firms' autarky payoffs to 1)

$$\pi_i(\mathbf{p}; I \times F) = \frac{1}{2}(\mathbf{a} - \mathbf{p})B^{-1}(\mathbf{a} - \mathbf{p}) + 1 \qquad \pi_f(\mathbf{p}; I \times F) = (p_f - c_f)D_f(\mathbf{p}) + 1$$
$$\pi_i(p_f; \{(i, f)\}) = \frac{1}{2B_{ff}}(a_f - p_f)^2 + 1 \qquad \pi_f(p_f; \{(i, f)\}) = (p_f - c_f)\hat{D}_f(p_f) + 1.$$

Given $\alpha \in \Delta(I \cup F)$, the first-order conditions for the Nash-in-Nash solution (5) with pairwise bargaining weights $\alpha_{if} = \frac{\alpha_i}{\alpha_i + \alpha_f}$ are

$$\alpha_f \frac{D_f(\mathbf{p}) + p_f B_{ff}^{-1}}{\pi_f(\mathbf{p}; I \times F)} - \alpha_i \frac{D_f(\mathbf{p})}{\pi_i(\mathbf{p}; I \times F) - \pi_i(p_{-f}; \{(i, -f)\})} = 0, \ f \in \{1, 2\}.$$

For the conditions that pin down the outcome that is stable given the beliefs associated with equal bargaining weights, first observe that the nonstrategic individual rationality constraints in (4) cannot bind for any of the firms, regardless of the set of contracting firms A. Assuming there is no inefficient exclusion, the first-order conditions for the solution to (4) are thus given by

$$\alpha_f \frac{D_f(\mathbf{p}) + p_f B_{ff}^{-1}}{\pi_f(\mathbf{p}; I \times F)} + \alpha_f \frac{p_{-f} B_{12}^{-1}}{\pi_{-f}(\mathbf{p}; I \times F)} - \alpha_i \frac{D_f(\mathbf{p})}{\pi_i(\mathbf{p}; I \times F)} = 0, \ f \in \{1, 2\}.$$

Unlike the Nash-in-Nash FOCs, these conditions capture the way that the downstream firm takes into account the way that an increase in the price p_f charged by one supplier increases the surplus available between the downstream firm and the other supplier, by adding a positive (since B^{-1} has nonnegative entries[10]) term to the derivative that must equal zero.

Finally, unlike (5), no "outside option" terms appear in the objective function in (4). This is because firms' outside options to a contract (or set of contracts) are already explicitly considered by the underlying stability concept. (4) is just a "reduced form" for stability with correct, consistent beliefs that is provided by Proposition 1. As a result of Proposition 1, the impact of firms' outside options on (4) is accounted for through the nonstrategic individual rationality constraints (NSIR(i)) and (NSIR(f)).

[10]This follows from the fact that B is an M-matrix.

Comparison to Nonstrategic Stability-Based Comparative Statics

It is helpful to contrast our approach with the comparative statics that are possible using the standard nonstrategic approach to matching-theoretic stability. First, in many settings of interest — such as those where agreements have externalities, agents have complex preferences, or agreements can involve more than two agents — the nonstrategic approach may fail to make a prediction in a counterfactual scenario, even if we know the agents' preferences: the set of stable outcomes may be empty. But even in environments where the canonical approach to matching-theoretic stability does not face existence issues, using it for counterfactual analysis is not always feasible.

For instance, suppose that one or more contracts is rendered unavailable (e.g., through regulation). If we can identify agents' preferences, we can use the standard, nonstrategic approach to stability to compute a new set of stable outcomes after contracts are removed from the environment. However, even if this set is not empty, it may contain more than one outcome. If it does, we do not know which of the new stable outcomes will result, because that depends on the sequence of deviations that follows the contract's removal. In contrast, the approach based on strategic consistency that we outline in this paper pins down a unique stable outcome that is independent of the path taken to get to it: Since agents' beliefs are cross-set consistent, any sequence of deviations following the contract's removal must lead to the same outcome.[11]

More generally, the map between strategically consistent profiles (and hence stable outcomes) in different environments provided by Proposition 1 allows us to uniquely pin down the impact on beliefs and outcomes of arbitrary changes to the environment: e.g., changes to the set of contracts X (e.g., replacing intermediated trades with direct ones), changes to the set of agents $N(x)$ named by a contract (e.g., giving a regulator the ability to veto it), or changes to the agents' payoff functions u_j (e.g., through a merger, the imposition of a tax, or the introduction of other types of externalities).

[11] In particular, we can simply restrict the domain of beliefs and choice functions to sets that do not include the removed contracts, and recover the new outcome from (SPP(Y)) by letting Y be the set of remaining contracts.

4 Conclusion

This paper shows how matching-theoretic stability can be used for counterfactual analysis in the same way (and in the same environments) as the Nash-in-Nash solution concept is commonly used. This allows an analyst to sidestep the existence issues faced by stability in environments with externalities without restricting the alternative sets of agreements that agents can consider. It also allows for endogenous exclusion.

References

BAGWELL, K., R. W. STAIGER, AND A. YURUKOGLU (2021): "Quantitative Analysis of Multiparty Tariff Negotiations," *Econometrica*, 89, 1595–1631.

CLARK, R., I. HORSTMANN, AND J.-F. HOUDE (2024): "Hub-and-Spoke Cartels: Theory and Evidence From the Grocery Industry," *American Economic Review*, 114, 783–814.

COLLARD-WEXLER, A., G. GOWRISANKARAN, AND R. S. LEE (2019): ""Nash-in-Nash" Bargaining: A Microfoundation for Applied Work," *Journal of Political Economy*, 127, 163–195.

CRAWFORD, G. S. AND A. YURUKOGLU (2012): "The Welfare Effects of Bundling in Multichannel Television Markets," *American Economic Review*, 102, 643–85.

DRANOVE, D. (1998): "Economies of Scale in Non-Revenue Producing Cost Centers: Implications for Hospital Mergers," *Journal of health economics*, 17, 69–83.

GRENNAN, M. (2013): "Price Discrimination and Bargaining: Empirical Evidence from Medical Devices," *American Economic Review*, 103, 145–77.

HO, K. AND R. S. LEE (2017): "Insurer Competition in Health Care Markets," *Econometrica*, 85, 379–417.

——— (2019): "Equilibrium Provider Networks: Bargaining and Exclusion in Health Care Markets," *American Economic Review*, 109, 473–522.

HORN, H. AND A. WOLINSKY (1988): "Bilateral monopolies and incentives for merger," *The RAND Journal of Economics*, 408–419.

NASH, J. F. (1950): "The Bargaining Problem," *Econometrica: Journal of the Econometric Society*, 155–162.

ROSTEK, M. AND N. YODER (2025): "Matching with Strategic Consistency," Working Paper.

A Proof of Proposition 1

Label $I = \{1, 2, \ldots, |I|\}$ and $F = \{1, 2, \ldots, |F|\}$.

Part 1: Construction of $\{C_j^{\phi}, \mu_j^{\phi}\}_{j \in I \cup F}$

First, for all sets $Y \subseteq X$ that are not compact, let $\mu_j^{\phi}(Y) = \varnothing$ and $C_j^{\phi}(Y_j | Y_{-j}) = \varnothing$ for each $j \in I \cup F$. Then for each j, $\mu_j^{\phi}(Y)$ and $C_j^{\phi}(Y_j | Y_{-j})$ satisfy (1) and (2).

Now suppose that $Y \subseteq X$ is compact, and so $\{p_{if} | (i, f, p_{if}) \in Y\}$ is compact for each $(i, f) \in I \times F$.

Existence of solutions to the social planner's problem. By assumption, if $u_k(S) \geqslant u_k(S' \cup S_{-k})$ for all $k \in I \cup F$ and $S' \subset S_k$, then there do not exist (i, f) and distinct p_{if}, p'_{if} such that $(i, f, p_{if}) \in S$ and $(i, f, p'_{if}) \in S$. Then (SPP(Y)) can be written

$$\mu_j^{\phi}(Y) = \{(i, f, p_{if})\}_{(i,f) \in A}$$

for some $(A, \mathbf{p}_A) \in M^*(Y) = \arg \max_{\substack{A \subseteq I \times F \\ \mathbf{p}_A \in \times_{(i,f) \in A} \{p_{if} | (i, f, p_{if}) \in A\}}} \phi((\pi_j(\mathbf{p}_A; A))_{j \in I \cup F})$

$$\tag{SPP$'$(Y)}$$

$$\text{s.t. } \pi_i(\mathbf{p}_A; A) \geqslant \pi_i(\mathbf{p}_{A \setminus B}; A \setminus B) \text{ for all } i \in I \text{ and } B \subseteq \{i\} \times F; \tag{NSIR(i)}$$

$$\pi_f(\mathbf{p}_A; A) \geqslant \pi_f(\mathbf{p}_{A \setminus B}; A \setminus B) \text{ for all } f \in F \text{ and } B \subseteq \{f\} \times I. \tag{NSIR(f)}$$

Since each π_j is continuous and I and F are finite, for each $A \subseteq I \times F$, the set

$$P_A^{IR} = \left\{ \mathbf{p}_A \;\middle|\; \begin{array}{c} \pi_i(\mathbf{p}_A; A) \geqslant \pi_i(\mathbf{p}_{A \setminus B}; A \setminus B) \text{ for all } i \in I \text{ and } B \subseteq \{i\} \times F \\ \pi_f(\mathbf{p}_A; A) \geqslant \pi_f(\mathbf{p}_{A \setminus B}; A \setminus B) \text{ for all } f \in F \text{ and } B \subseteq \{f\} \times I \end{array} \right\}$$

is closed. Then $\bigcup_{A \in 2^{I \times F}} \left(\{A\} \times \left(P_A^{IR} \cap \times_{(i,f) \in A} \{p_{if} | (i, f, p_{if}) \in Y\} \right) \right)$ is compact. Since each π_j is continuous, and ϕ is continuous, so is $\phi \circ (\pi_j)_{j \in I \cup F}$; it follows from the Maximum Theorem that the set $M^*(Y)$ is nonempty and compact.

Consistent tiebreaking in the social planner's problem. Define the order $\succeq_{I \times F}$ on $2^{I \times F}$ as follows: $A \succ_{I \times F} B$ if for some $(i, f) \in I \times F$, $(i, f) \in A \setminus B$, $\nexists (i', f') \in B \setminus A$ with

$i' < i$ or $i' = i$ and $f' < f$. Clearly, $\geq_{I\times F}$ is a total order, and if $A \supset B$, then $A >_{I\times F} B$.

Let $A^*(Y) = \max_{\geq_{I\times F}}\{A|(A,\mathbf{p}_A) \in M^*(Y)\}$. Clearly, if $M^*(Y) \supseteq M^*(Z)$, then $A^*(Y) \geq_{I\times F} A^*(Z)$. Thus, if $(A^*(Y), \mathbf{p}_{A^*(Y)}) \in M^*(Z)$ for some $\mathbf{p}_{A^*(Y)}$, then $A^*(Y) = A^*(Z)$.

Define $\mathbf{p}^*_{A^*(Y)}(Y) \in \mathcal{P}^{A^*(Y)}$ recursively as follows:

Step $(1,1)$**.** Define

$$
\begin{aligned}
p^*_{11}(Y) &= \min\{p_{11}|(A^*(Y),\mathbf{p}_A) \in M^*(Y)\} \text{ if } (1,1) \in A^*(Y) \\
P_{1,1}(Y) &= \begin{cases} \{A^*(Y)\} \times \mathbb{R}_+^{A^*(Y)} & (1,1) \notin A^*(Y) \\ \{A^*(Y)\} \times \mathbb{R}_+^{A^*(Y)\setminus\{(1,1)\}} \times \{p^*_{11}(Y)\} & (1,1) \in A^*(Y) \end{cases}
\end{aligned}
$$

Note that p^*_{11} is well defined since it is the minimum of a projection of the compact set $M^*(Y) \cap \left(\{A^*(Y)\} \times \mathbb{R}_+^{A^*(Y)}\right)$, and that $P_{1,1}(Y)$ is compact.

Step $(1,f)$**.** For each $f > 1$, recursively define

$$
\begin{aligned}
p^*_{1f} &= \min\{p_{1f}|(A^*(Y),\mathbf{p}_A) \in M^*(Y) \cap P_{1,f-1}\} \text{ if } (1,f) \in A^*(Y) \\
P_{1,f}(Y) &= \begin{cases} P_{1,f-1}(Y) & (1,f) \notin A^*(Y) \\ P_{1,f-1}(Y) \cap \left(\{A^*(Y)\} \times \mathbb{R}_+^{A^*(Y)\setminus\{(1,f)\}} \times \{p^*_{1f}(Y)\}\right) & (1,f) \in A^*(Y) \end{cases}
\end{aligned}
$$

Note that each p^*_{1f} is well defined since it is the minimum of a projection of the compact set $M^*(Y) \cap P_{1,f-1}(Y)$, and that each $P_{1,f}(Y)$ is compact.

Then, for each $i > 1$, proceed recursively as follows: **Step** $i,1$**.** Define

$$
\begin{aligned}
p^*_{i1}(Y) &= \min\{p_{i1}|(A^*(Y),\mathbf{p}_A) \in M^*(Y) \cap P_{i-1,|F|}\} \text{ if } (i,1) \in A^*(Y) \\
P_{i,1}(Y) &= \begin{cases} P_{i-1,|F|}(Y) & (i,1) \notin A^*(Y) \\ P_{i-1,|F|}(Y) \cap \left(\{A^*(Y)\} \times \mathbb{R}_+^{A^*(Y)\setminus\{(i,1)\}} \times \{p^*_{i1}(Y)\}\right) & (i,1) \in A^*(Y) \end{cases}
\end{aligned}
$$

Note that each p^*_{i1} is well defined since it is the minimum of a projection of the compact set $M^*(Y) \cap P_{i-1,|F|}$, and that each $P_{i,1}(Y)$ is compact.

Step (i, f)**.** For each $f > 1$, recursively define

$$p^*_{if}(Y) = \min\{p_{if} | (A^*(Y), \mathbf{p}_A) \in M^*(Y) \cap P_{i,f-1}\} \text{ if } (i, f) \in A^*(Y)$$
$$P_{i,f}(Y) = \begin{cases} P_{i,f-1}(Y) & (1, f) \notin A^*(Y) \\ P_{i,f-1}(Y) \cap \left(\{A^*(Y)\} \times \mathbb{R}_+^{A^*(Y)\backslash\{(i,f)\}} \times \{p^*_{if}(Y)\}\right) & (i, f) \in A^*(Y) \end{cases}$$

Note that each p^*_{if} is well defined since it is the minimum of a projection of the compact set $M^*(Y) \cap P_{i,f-1}$, and that each $P_{i,f}(Y)$ is compact.

Finally, note that $P_{|F|,|F|} = p^*_{A^*(Y)}(Y)$, and so $(A^*(Y), p^*_{A^*(Y)}) \in M^*(Y)$. Furthermore, observe that since $p^*_{A^*(Y)}$ is defined by recursively taking pointwise minima, if $M^*(Y) \supseteq M^*(Z)$ and $(A^*(Y), p^*_{A^*(Y)}) \in M^*(Z)$, then since $A^*(Y) = A^*(Z)$, we must have $(A^*(Y), p^*_{A^*(Y)}) = (A^*(Z), p^*_{A^*(Z)})$.

Beliefs and Choices. Let $\mu^\phi_j(Y) = \{(i, f, p^*_{if}(Y))\}_{(i,f)\in A^*(Y)}$, and $C^\phi_j(Y_j|Y_{-j}) = \mu^\phi_j(Y) \cap X_j$. Then by definition, $\mu^\phi_j(Y)$ satisfies (1). Since $(A^*(Y), p^*_{A^*(Y)})$ satisfies (NSIR(f)) and (NSIR(i)), $C^\phi_j(Y_j|Y_{-j})$ satisfies (2), and moreover, $\mu^\phi_j(Y)$ solves (SPP(Y)).

Part 2: Strategic Consistency of $\{C^\phi_j, \mu^\phi_j\}_{j\in I\cup F}$

It follows immediately from the previous step that each μ^ϕ_j is correct given $\{C^\phi_k\}_{k\neq j}$, and each C_j is optimal given μ^ϕ_j. It remains to show cross-set consistency.

Suppose that $Y, Z \subseteq X$ are compact, and that $Y \supseteq Z \supseteq C^\phi_j(Y_j|Y_{-j})$ for all $j \in I \cup F$. By definition, $C^\phi_j(Y_j|Y_{-j}) = \{(j, f, p^*_{jf}(Y))\}_{(j,f)\in A^*(Y)}$ for each $j \in I$ and $C^\phi_j(Y_j|Y_{-j}) = \{(i, j, p^*_{ij}(Y))\}_{(i,j)\in A^*(Y)}$ for each $j \in F$. Hence, if $Y \supseteq Z \supseteq C^\phi_j(Y_j|Y_{-j})$ for all $j \in I \cup F$, we must have $\{(i, f, p^*_{if}(Y))\}_{(i,f)\in A^*(Y)} \in Z$. Then since $(A^*(Y), p^*_{A^*(Y)})$ solves (SPP'(Y)), (SPP(Z)) must attain the same value, and hence we must have $(A^*(Y), p^*_{A^*(Y)}) \in M^*(Z)$ and $M^*(Y) \supseteq M^*(Z)$. As we showed in our construction of $(A^*(Y), p^*_{A^*(Y)})$, this implies that $(A^*(Y), p^*_{A^*(Y)}) = (A^*(Z), p^*_{A^*(Z)})$. Then by definition, $\mu^\phi_j(Y) = \mu^\phi_j(Z)$ for all $j \in I \cup F$, as desired.

Part 3: Pareto Optimality of $\{C^\phi_j, \mu^\phi_j\}_{j\in I\cup F}$

Suppose Y and Z are nonstrategically individually rational, and that Z Pareto improves upon Y. Since any pair of agents signs at most one contract in any NSIR outcome, $Y \cup Z$ is finite, and hence compact. Then by construction, $\mu_j^{\phi}(Y \cup Z)$ solves (SPP($Y \cup Z$)). But Y cannot be a solution to this problem, because Z Pareto improves upon it and ϕ is strictly increasing.

Part 4: $\{(i, f, p^*_{if}(X))\}_{(i,f) \in A^*(X)}$ is uniquely stable given $\{C_j^{\phi}\}_{j \in I \cup F}$

By cross-set consistency, for each $j \in I \cup F$, $\{(i, f, p^*_{if}(X))\}_{(i,f) \in A^*(X)} \cap X_j = \mu_j^{\phi}(X) \cap X_j = C_j(Z_j | Z_{-j})$ for any $Z \supseteq \mu_j(X) = \{(i, f, p^*_{if}(X))\}_{(i,f) \in A^*(X)}$. It follows that $\{(i, f, p^*_{if}(X))\}_{(i,f) \in A^*(X)}$ is individually rational and unblocked. □