Dynamic Oligopoly Pricing with Asymmetric Information: Implications for Horizontal Mergers*

Andrew Sweeting†
University of Maryland, CEPR and NBER

Xuezhen Tao
Shanghai University of Finance and Economics

Xinlu Yao

June 2021

Abstract

We model differentiated product pricing by firms that possess private information about serially-correlated state variables, such as their marginal costs, and can use prices to signal information to rivals. In a dynamic game, signaling can raise prices significantly above static complete information Nash levels even when the privately observed state variables are restricted to lie in narrow ranges. We calibrate our model using data from the beer industry, and we show that our model can explain changes in price levels and price dynamics after the 2008 MillerCoors joint venture.

JEL CODES: D43, D82, L13, L41, L93.

Keywords: oligopoly, dynamic games, dynamic pricing, asymmetric information, signaling, firm conduct, tacit collusion, horizontal mergers, merger retrospectives, beer industry.

---

*We thank Nate Miller, Matt Weinberg, Gloria Sheu, Dan Vincent, Bob Majure, Joe Harrington, Kyle Bagwell and a number of seminar participants for valuable comments, which have been especially useful in framing the paper. This paper uses the IRI Academic Dataset (Bronnenberg, Kruger, and Mela (2008)). All estimates and analysis using IRI data are the work of the authors and not IRI. This research was supported by US National Science Foundation Grant SES-1658670 (Sweeting) and National Natural Science Foundation of China Grant 72003120 (Tao). Xinlu Yao's work on this project was completed when she was a graduate student at the University of Maryland. All errors are our own.

†Corresponding author. Contact: Department of Economics, 3135 Tydings Hall, University of Maryland, College Park, MD 20742, USA. Email: atsweet@umd.edu.

1 Introduction

Theoretical and empirical analyses of supply in differentiated product markets usually assume that firms have complete information (CI) and set prices to maximize their current profits. If an alternative is considered, it is typically tacit collusion with repeated CI stage games, which, in the empirical literature is often modeled using a "conduct parameter" (Bresnahan (1982), Lau (1982), Nevo (1998)) where each firm uses a standard CI Nash first-order conditions except that some weight is placed on the profits of its rivals. These CI formulations are tractable and, under appropriate assumptions, they are econometrically identified (Berry and Haile (2014)).

However, assuming that firms have CI about all factors that may affect their rivals' pricing choices is a strong assumption. Public companies closely guard information about the profitability of individual product lines and government agencies presume that information on revenues, costs and margins is competitively sensitive and highly confidential during antitrust investigations even while they use models that assume CI to model market outcomes. There is also surprisingly little evidence that CI oligopoly models accurately predict qualitative or quantitative changes in pricing behavior after structural changes in market conditions, such as the consummation of a merger.

It is clearly important to know whether predictions would change, in a material way, if the CI assumption is relaxed. A natural assumption for an economist is that, when firms have privately-observed state variables, a firm will try to learn from its rivals' choices (i.e., their prices) about those variables, in order to try to more accurately predict how those rivals will price in the future. If this happens, then firms may also have incentives to distort their prices in order to affect what their rivals will expect.

We develop models where this logic applies. Specifically, we will assume that each firm has a payoff-relevant state variable, such as its marginal cost, which is positively but imperfectly serially-correlated and unobserved by rivals. Prices are perfectly observed. We will consider fully separating equilibria where, in equilibrium, a firm's chosen price perfectly reveals its current cost, and beliefs have a simple form. In these equilibria, all firms that do not have the lowest possible marginal cost set prices above static best response levels to credibly signal this information to their rivals. This can, in turn, cause static best response prices to increase, and signaling prices to rise further, a positive feedback that can cause equilibrium prices to be significantly above static CI Nash levels, although, as we discuss, separating equilibria may not exist if prices rise too much. The model also explains significant equilibrium volatility in prices even when the underlying range over which the privately-observed state variables can vary is very limited. We provide examples where private information about 5 cents, or less than 1%, of marginal costs, which an empirical economist might be tempted to ignore, can raise average prices, relative to a static CI model, by more than a couple of dollars or more than 10%, while increasing the standard deviation of prices by a factor of 40. While a small theoretical literature has shown that oligopoly signaling can affect equilibrium prices in two- or three-period models, we provide the first analysis of how large these effects may be, and the first empirical application.

We apply our model to horizontal merger analysis, as CI Nash first-order conditions are relied upon by agencies when deciding whether to challenge a proposed transaction. Signaling is a strategic investment to raise rivals' future prices, and like many strategic investments, the equilibrium incentive to invest can rise when the number of competitors is reduced. We use examples to illustrate that standard static CI merger simulations can significantly underpredict how much both merging parties and their rivals increase prices, or, equivalently, how large marginal cost synergies have to be to prevent prices from rising. These biases exist when firms are symmetric and when there are asymmetries pre-merger.

Our empirical application calibrates our model to data from the U.S. beer market around the time of the 2008 Miller-Coors (MC) joint venture (JV). Miller and Weinberg (2017) (MW) show that, after the JV, MC and its larger domestic rival Anheuser-Busch (AB), increased prices in a way that is inconsistent with static CI Nash pricing, and they estimate a conduct parameter that rationalizes the average price increases assuming CI. Miller, Sheu, and Weinberg (2020) (MSW) propose a more complete model of tacit collusion under CI, where domestic brewers set common supermarkups in each regional market, and they estimate a time preference parameter to rationalize the size of the post-JV price increase. We take a different approach by calibrating our dynamic signaling model to match price dynamics in pre-JV data. Without introducing any additional parameters, we show that our model accurately predicts the observed increase in average prices and that it also correctly predicts directional changes in which several statistics that summarize price dynamics.[1] On the other hand, a calibrated CI model fits the pre-JV

[1]Appendix D.9 shows that the assumptions that MW and MSW make on domestic brewer behavior can also

data less well, and, even when we allow for a post-JV conduct parameter to explain the change in average prices, it fails to predict how price dynamics change.[2]

Before discussing the related literature, we should be clear about several limitations of our analysis, which reflect the novelty of the quantitative exercises we conduct. First, we have to assume that each firm has exactly one privately-known state variable and can send exactly one signal per period. This imposes restrictions on the types of mergers that can plausibly be considered, although we will argue that the assumptions are reasonable in our application. Second, we only consider fully separating equilibria, even though these may not exist for some parameters and we can only prove existence and uniqueness in special cases. It is possible that pooling equilibria could generate substantially larger price effects. Third, folk theorems imply that collusive conduct under CI could take many forms, some of which might generate similar changes in price dynamics that our model can generate. Therefore, while we present evidence against particular collusive models (Appendix D.9), it is not possible to generally reject tacit collusion as a possible explanation for changes in pricing after a merger. However, our results do show that tacit collusion or market features that are often considered as pre-requisites of collusion, such as transparency of costs or firm symmetry, are not required to explain why static CI Nash pricing models fail to predict accurately.

The rest of this introduction reviews the related literature. Section 2 lays out the model and the equilibrium concept. Section 3 presents some examples and illustrates the implications for merger analysis. Section 4 provides our empirical application. Section 5 concludes. The Appendices, intended for online publication, detail the computational algorithms, additional examples, a proof of existence and uniqueness for the case of linear demand, and details of the data and additional empirical analyses.

Related Literature. Shapiro (1986) and Vives (2011) examine how equilibrium prices and welfare change when marginal costs are private information in one-shot oligopoly models. Most of our focus will be on models where marginal costs lie in quite narrow intervals and the static effects that these papers identify are very small. A large theoretical literature has considered

be rejected using alternative exclusion restrictions and richer fixed effects, which suggests that one should be reluctant to use these models to prospectively predict the effects of mergers.

[2]Our analysis of price dynamics, and what it implies for firm behavior, is also new to the literature analyzing horizontal mergers, even though prices are quite volatile in many industries, such as airlines, where mergers have been extensively studied.

one-shot signaling models where only one player has private information. The classic Industrial Organization example is the Milgrom and Roberts (1982) limit pricing model, where an incumbent monopolist may lower its first period price to deter entry in a two-period game. Sweeting, Roberts, and Gedge (2020) develop finite and infinite-horizon versions of this model where an incumbent monopolist's type changes over time, as we will assume in this paper.[3] They estimate the model and show that it can explain why incumbent airlines dropped prices by as much as 15% when Southwest threatened entry on monopoly routes. The model with several incumbents considered here is potentially applicable to a much wider range of industries.

The literature on games where multiple players signal simultaneously is limited.[4] Mailath (1988) identifies conditions under which a separating equilibrium will exist in an abstract two-period game with continuous types, and shows that the conditions on payoffs required for the uniqueness of each player's separating best response function are similar to those shown by Mailath (1987) for models where only one player is signaling (Mailath and von Thadden (2013) generalize these conditions). Mailath (1989) applies these results to a two-period pricing game where differentiated firms have static linear demands and marginal costs that are private information but fixed. Firms raise their prices in the first period in order to try to raise their rivals' prices in the second period.[5] Mester (1992) extends this approach to a three-period quantity-setting model where marginal costs change over time, and she shows that signaling, which leads to increased output in this case, happens in the first two periods.

We rely on Mailath's results to characterize best response signaling pricing functions, and we will focus on the magnitude, empirical relevance and implications of the equilibrium effects in multi-period settings with more standard forms of differentiated product demand. Fershtman and Pakes (2012) and Asker, Fershtman, Jeon, and Pakes (2020) develop an alternative approach to discrete state and discrete action dynamic games with asymmetric information. They reduce the computational burden using the concept of Experience-Based Equilibrium (EBE) where firms have beliefs about their payoffs from different actions rather than rivals'

[3]Kaya (2009) and Toxvaerd (2017) analyze one-sided, dynamic signaling games where the informed firm's type is fixed, and, in equilibrium, the informed firm signals until its reputation is established.

[4]Bonatti, Cisternas, and Toikka (2017) analyze linear signaling strategies in a continuous-time Cournot game where each firm's marginal cost is private information and fixed, but firms cannot perfectly observe the quantities that their rivals choose. We will assume that prices are perfectly observable.

[5]Caminal (1990) considers a two-period linear demand duopoly model where firms have private information about the demand for their own product, and also raise prices to signal that they will set higher prices in the final period.

types.[6] Our equilibrium concept is more standard, and the computational burden is reduced by focusing on fully separating equilibria in continuous action games.

The conclusion will discuss the relationship between our paper and discussions of coordinated effects in horizontal merger analysis (Ordover (2007), Baker and Farrell (2020), Farrell and Baker (2021)). Our paper is partly motivated by the empirical horizontal merger retrospectives literature that has often found that, presumably contrary to what agencies expected, prices often rise after mergers are consummated. Ashenfelter, Hosken, and Weinberg (2014) find that 36 of 49 studies across several industries identify significant post-merger price increases.[7] Peters (2009) and Garmon (2017) show that merger simulations and other methods, such as pricing pressure indices, that are derived from static CI first-order conditions often perform poorly at predicting price changes after airline and hospital mergers. This leads naturally to the question of which alternative models can do better.

2 Model

In this section, we present our model. More specific assumptions will be made in our examples and application.

2.1 Outline.

There are discrete time periods, $t = 1, ..., T$, where $T \leq \infty$, with discount factor $0 < \beta < 1$. $\beta = 0.99$, except when we show that our results are not particularly sensitive to the choice of discount factor. There are a fixed set of N risk-neutral firms. Each firm either sells a single-product or sells multiple products, which we will assume are symmetric in demand and are produced at the same marginal cost, at a single price. There may be observed and fixed differences in demand and costs across firms, but exactly one dimension of a firm's type is private information.

[6] The rest of the literature on dynamic games, following Ericson and Pakes (1995) and Pakes and McGuire (1994), has assumed that players observe all state variables apart from iid shocks to the payoffs from different actions, eliminating any role for signaling.

[7] Ashenfelter, Hosken, and Weinberg (2014) note that retrospectives have not typically found price increases in banking. Interestingly, the Mester (1992) analysis of a Cournot oligopoly model with asymmetric information was explicitly motivated by a desire to explain why, contrary to the predictions of Nash and tacit collusion models, concentration appeared to lead to more competitive behavior in banking.

We will consider two different formulations which we will use for different purposes. Our explanation of the model, our empirical application and the example that we use to provide intuition for how signaling affects prices will assume that the type is continuous on a known compact interval $[\underline{\theta_i}, \overline{\theta_i}]$. However, we use a model where firms can have two discrete types, $\underline{\theta_i}$ and $\overline{\theta_i}$ when we want to explore what will happen for many different parameters or different numbers of firms as the computational burden is lower. In both cases, types are assumed to evolve exogenously, and independently, from period-to-period according to a first-order Markov process, $\psi_i : \theta_{i,t-1} \rightarrow \theta_{i,t}$. This assumption is consistent with the treatment of productivity changes in the structural production function literature following Olley and Pakes (1996) (except Doraszelski and Jaumandreu (2013)).

2.2 Within-Period Timing.

In each period t of the game, timing is as follows. Firms enter period t with their $t-1$ types, which then evolve according to ψ_i. Firms observe their own new types, but neither the previous nor the new types of other firms.[8] Each firm then simultaneously chooses a price, $p_{i,t}$, with no menu costs. Once a firm sets its period t price, it is unable to change it. A firm's profits are given by $\pi_i(p_{i,t}, p_{-i,t}, \theta_{i,t})$ and we assume that $\frac{\partial \pi_i}{\partial p_{-i,t}} > 0$ for all $-i$. Note that $\pi_i(p_{i,t}, p_{-i,t}, \theta_{i,t})$ only depends on current prices and the firm's current type, consistent with static and time-invariant demand. Current and past prices are assumed to be perfectly observed by each firm.

2.3 Assumptions.

For continuous types, we make the following assumption.

Assumption 1 ***Type Transitions for the Continuous Type Model.*** *The conditional pdf $\psi_i(\theta_{i,t}|\theta_{i,t-1})$*

1. *has full support, so that the type can transition from any value on the support to any other value in a single period.*

2. *is continuous and differentiable (with appropriate one-sided derivatives at the boundaries).*

[8]Our fully separating equilibria would be unchanged if $t-2$ types were revealed.

3. for any $\theta_{i,t-1}$ there is some θ' such that $\frac{\partial \psi_i(\theta_{i,t}|\theta_{i,t-1})}{\partial \theta_{i,t-1}}|_{\theta_{i,t}=\theta'} = 0$ and $\frac{\partial \psi_i(\theta_{i,t}|\theta_{i,t-1})}{\partial \theta_{i,t-1}} < 0$ for all $\theta_{i,t} < \theta'$ and $\frac{\partial \psi_i(\theta_{i,t}|\theta_{i,t-1})}{\partial \theta_{i,t-1}} > 0$ for all $\theta_{i,t} > \theta'$.

This assumption implies types are positively, but not perfectly, serially correlated so that a higher type in one period implies that a higher type in the next period is more likely.

Beliefs about rivals' types play an important role in our game. In a fully separating equilibrium, each firm will (correctly) believe that each rival has a particular type in the previous period. For convenience, we assume that beliefs about types in $t = 1$ have the same structure.

Assumption 2 ***Initial Period Beliefs.*** *Firms know what their rivals' types were in a fictitious prior period, $t = 0$.*

2.4 Fully Separating Equilibrium in a Finite Horizon and Continuous Type Game.

We now describe the equilibrium for a game with two ex-ante symmetric single-product duopolists, which we will use in our first example.

2.4.1 Final Period (T).

In the final period, each firm maximizes its expected payoff given its own type, its beliefs about the its rival's type and its pricing strategy. Play is therefore consistent with a Bayesian Nash Equilibrium. If firm j believes that firm i's period $T-1$ type was $\widehat{\theta^j_{i,T-1}}$ and j's period T pricing function is $P_{j,T}(\theta_{j,T}, \theta_{j,T-1}, \widehat{\theta^j_{i,T-1}})$[9], then a type $\theta_{i,T}$ i will set a price

$$p^*_{i,T}(\theta_{i,T}, \theta_{j,T-1}, \widehat{\theta^j_{i,T-1}}) = \arg\max_{p_{i,T}} \int_{\underline{\theta_j}}^{\overline{\theta_j}} \pi(p_{i,T}, P_{j,T}(\theta_{j,T}, \theta_{j,T-1}, \widehat{\theta^j_{i,T-1}}), \theta_{i,T}) \psi(\theta_{j,T}|\theta_{j,T-1}) d\theta_{j,T}.$$

2.4.2 Earlier Periods ($1, .., T-1$).

In earlier periods, i may choose not to set a static best response price in order to affect j's belief about its type. The equilibrium concept that we use is symmetric Markov Perfect Bayesian Equilibrium (MPBE) (Toxvaerd (2008),Roddie (2012)). An MPBE specifies period-specific

[9]This notation reflects the fact that we are assuming that player j used an equilibrium strategy in $T-1$ that revealed its type ($\theta_{j,T-1}$), but we are allowing for the possibility that firm i may have deviated so that j's beliefs about i's previous type are incorrect.

pricing strategies for each firm i as a function of i's current type, i's belief about j's previous type, and j's belief about i's previous type. It also specifies each firm's belief about its rival's type given observed histories of prices. Equilibrium beliefs should be consistent with Bayes Rule given equilibrium pricing strategies. If there are multiple rivals, they should all have the same beliefs given an observed history. While only current types and prices are directly payoff-relevant, history can matter in this Markovian equilibrium because it affects beliefs. We will only consider fully separating MPBEs where, in every period, a firm's equilibrium pricing strategy perfectly reveals its current type, and j's belief about i's current type will come from inverting i's pricing function.

2.4.3 Characterization of Separating Pricing Functions in Period $t < T$.

We follow Mailath (1989), who shows that one can apply the results in Mailath (1987) to this problem, in characterizing fully separating pricing functions using a definition of firm i's period-specific "signaling payoff function", $\Pi^{i,t}(\theta_{i,t}, \widehat{\theta^j_{i,t}}, p_{i,t})$. This is the present discounted value of firm i's expected current and future payoffs when its current type is $\theta_{i,t}$, it sets price $p_{i,t}$ and j believes, *at the end of period* t, that i has type $\widehat{\theta^j_{i,t}}$. $\Pi^{i,t}$ is assumed to be continuous and at least twice differentiable in its arguments. It is implicitly conditional on (i) j's period t pricing strategy, which will depend on j's beliefs about $t-1$ types, and (ii) both players' strategies in future periods. As j's end-of-period t belief about i's type enters as a separate argument, $p_{i,t}$ only affects $\Pi^{i,t}$ through period t profits. Given conditions on $\Pi^{i,t}$ that will be listed in a moment, the fully separating best response function of firm i, which is also implicitly conditioned on j's current pricing strategy and beliefs about previous types, can be uniquely characterized as follows (see Appendix C for a restatement of the Mailath (1987) theorems): i's pricing function will be the solution to a differential equation where

$$\frac{\partial p^*_{i,t}(\theta_{i,t})}{\partial \theta_{i,t}} = -\frac{\Pi^{i,t}_2\left(\theta_{i,t}, \widehat{\theta^j_{i,t}}, p_{i,t}\right)}{\Pi^{i,t}_3\left(\theta_{i,t}, \widehat{\theta^j_{i,t}}, p_{i,t}\right)} > 0, \tag{1}$$

and a boundary condition. The subscript n in $\Pi^{i,t}_n$ denotes the partial derivative of $\Pi^{i,t}$ with respect to the n^{th} argument. Assuming that lower types want to set lower prices (e.g., a type corresponds to the firm's marginal cost), the boundary condition will be that $p^*_{i,t}(\underline{\theta_i})$ is the solution to

$$\Pi_3^{i,t}\left(\underline{\theta_i}, \widehat{\theta_{i,t}^j}, p_{i,t}\right) = 0, \tag{2}$$

i.e., the lowest type's price maximizes its static expected profits given j's pricing policy. The numerator in (1) is i's marginal future benefit from raising j's belief about $\theta_{i,t}$, and the denominator is the marginal effect of a price increase on i's current profit. For prices above a static best response price, the denominator will be negative, and the pricing function will slope upwards in the firm's type.

This characterization of a separating best response will be valid under four conditions on $\Pi^{i,t}$, in addition to continuity and differentiability,

Condition 1 ***Shape of*** $\Pi^{i,t}$ ***with respect to*** $p_{i,t}$***.*** *For any* $(\theta_{i,t}, \widehat{\theta_{i,t}^j})$*,* $\Pi^{i,t}\left(\theta_{i,t}, \widehat{\theta_{i,t}^j}, p_{i,t}\right)$ *has a unique optimum in* $p_{i,t}$*, and, for all* $\theta_{i,t}$*, for any* $p_{i,t}$ *where* $\Pi_{33}^{i,t}\left(\theta_{i,t}, \widehat{\theta_{i,t}^j}, p_{i,t}\right) > 0$*, there is some* $k > 0$ *such that* $\left|\Pi_3^{i,t}\left(\theta_{i,t}, \widehat{\theta_{i,t}^j}, p_{i,t}\right)\right| > k$*.*

Condition 2 ***Type Monotonicity.*** $\Pi_{13}^{i,t}\left(\theta_{i,t}, \widehat{\theta_{i,t}^j}, p_{i,t}\right) \neq 0$ *for all* $(\theta_{i,t}, \widehat{\theta_{i,t}^j}, p_{i,t})$*.*

Condition 3 ***Belief Monotonicity.*** $\Pi_2^{i,t}\left(\theta_{i,t}, \widehat{\theta_{i,t}^j}, p_{i,t}\right)$ *is either* > 0 *for all* $(\theta_{i,t}, \widehat{\theta_{i,t}^j})$ *or* < 0 *for all* $(\theta_{i,t}, \widehat{\theta_{i,t}^j})$*.*

Condition 4 ***Single-Crossing.*** $\dfrac{\Pi_3^{i,t}\left(\theta_{i,t}, \widehat{\theta_{i,t}^j}, p_{i,t}\right)}{\Pi_2^{i,t}\left(\theta_{i,t}, \widehat{\theta_{i,t}^j}, p_{i,t}\right)}$ *is a monotone function of* $\theta_{i,t}$ *for all* $\widehat{\theta_{i,t}^j}$ *and for* $(\theta_{i,t}, p_{i,t})$ *in the graph of* $p_{i,t}^*(\theta_{i,t}, \theta_{j,t-1})$*.*

To interpret these conditions, assume that types correspond to marginal costs. The first condition will be satisfied if, for any marginal cost and distribution of prices that the rival may set, a firm's expected current period profit is quasi-concave in its own price. This will hold for common forms of differentiated product demand such as the multinomial and nested logit models. Type monotonicity requires that, when a firm increases its price, the profit that it loses will be lower if it has higher marginal costs. This will hold for constant marginal costs. Belief monotonicity requires that a firm's expected future profits should increase when rivals believe that it has a higher cost, holding its actual cost fixed. This condition may fail, and Appendix B.1.2 analyzes an example of failure in a two-type model. Single-crossing requires that a firm with a higher marginal cost should always be more willing to raise its price, reducing its current profits, in order to raise its rivals' beliefs about its marginal cost. This condition can also fail.

For completeness, we also need to define beliefs that a firm will have if the rival sets a price that is outside the range of the pricing function (i.e., a price that is not on the equilibrium path). When types correspond to marginal costs, we will assume that when a firm sets a price below (above) the lowest (highest) price in the range of the pricing function, it will be inferred to have the lowest (highest) possible cost type.

2.4.4 Existence and Uniqueness of a Fully Separating Equilibrium.

The conditions defined above guarantee the existence and uniqueness of fully separating best responses in any period, but this does not prove the existence or uniqueness of a fully separating equilibrium in the whole game. Mailath (1989) proves existence and uniqueness in a two-period duopoly game with linear demand and there is private information about marginal costs. Appendix C proves existence and uniqueness in a finite horizon, linear demand duopoly game where marginal costs are private information. The proof requires that the marginal cost interval $(\overline{\theta} - \underline{\theta})$ is small enough so that a single-crossing condition holds when prices rise. Of course, in the infinite horizon case, a CI pricing game has many equilibria.

In the continuous type case, if the conditions on payoff functions are satisfied, a firm will have a unique separating best response function (Mailath (1987)) given the strategies of the other firms. This is not the case in the two-type model, where it is possible to construct different separating best responses depending on the beliefs of the other firms. Therefore, in the two-type case, we need to use a refinement even to find best responses. Specifically, we solve for the strategies that achieve separation at the lowest cost to the signaling firm, consistent with the type of "intuitive criterion" (Cho and Kreps (1987)) refinement that has been widely used in one-sided signaling models with two types.

In our application, we will assume nonlinear demand and, to reduce the computational burden, an infinite horizon. We will therefore proceed without proofs of existence or uniqueness. Appendix A details how we compute equilibrium strategies, and verify belief monotonicity and single-crossing as part of the algorithm. We will discuss examples where we cannot find a separating equilibrium below. We have only ever found a single equilibrium in finite horizon games and infinite horizon games with continuous types, but we have found examples of multiplicity in infinite horizon games with two types even when we use the refinement to define best response functions.[10]

3 Examples

This section uses examples to illustrate equilibrium strategies and outcomes, and how the effects of signaling (relative to complete information) and the effects of mergers vary with the number of firms, asymmetries and the discount rate. We also summarize the results of some additional examples included in the Appendices.

3.1 Illustration of Equilibrium Strategies Using a Finite Horizon Continuous-Type Duopoly Example.

Specification. There are two ex-ante symmetric single-product firms. Demand is determined by a nested logit model, with both products in one nest, and the outside good in its own nest. Consumer c's indirect utility from buying from product i is $u_{i,c} = 5 - 0.1p_i + \sigma\nu_c + (1-\sigma)\varepsilon_{i,c}$ where p_i is the dollar price, $\varepsilon_{i,c}$ is a draw from a Type I extreme value distribution, $\sigma = 0.25$, and ν_c is an appropriately distributed draw for c's nest preferences. For the outside good, $u_{0,c} = \varepsilon_{0,c}$. We will set market size equal to 1, so that our welfare numbers have a "per-potential consumer" interpretation. We examine what happens to strategies in a finite horizon game with $T = 25$ periods as we believe that this helps to clarify intuitions. The game is solved backwards from the last period.

We assume that marginal cost is private information, and that, for each firm, it lies in the interval $[\underline{c}, \overline{c}] = [\$8, \$8.05]$. Costs evolve independently according to an exogenous truncated AR(1) process

$$c_{i,t} = \rho c_{i,t-1} + (1-\rho)\frac{\overline{c}+\underline{c}}{2} + \eta_{i,t}, \tag{3}$$

where $\rho = 0.8$ and $\eta_{i,t} \sim TRN(0, \sigma_c^2, \underline{c} - \rho c_{i,t-1} - (1-\rho)\frac{\overline{c}+\underline{c}}{2}, \overline{c} - \rho c_{i,t-1} - (1-\rho)\frac{\overline{c}+\underline{c}}{2})$. TRN denotes a truncated normal distribution. Its arguments are, in order, the mean and the variance

[10] In examples where we have found multiplicity, the algorithm that we use elsewhere in the paper appears to consistently pick out an equilibrium that is the limit of the equilibrium in the early periods of a finite horizon game as the number of periods grows.

of the untruncated distribution, and the lower and upper truncation points. $\sigma_c = \$0.025$.

Three features of this parameterization are worth highlighting. First, marginal costs are restricted to a narrow range (diverging by less than 0.32% from mean value) and the probability that a firm will switch from a relatively high cost to a relatively low cost across periods is quite high.[11] Therefore, no signal should affect a rival's posterior belief about a firm's next period marginal cost very much. These choices are deliberate, as, first, we want to emphasize that we find large effects of signaling on prices when the supply-side differs only slightly from what would be assumed in a CI analysis, and, second, as we will show, the support of marginal costs has to be relatively constrained for us to be able to find a separating equilibrium. Second, the demand parameters imply high margins and limited substitution to the outside good in both static and dynamic equilibria. We will also discuss how these features contribute to the existence of a fully separating equilibrium with large price effects. Finally, we are assuming that the firms are (ex-ante) symmetric. Later examples will analyze what happens when we allow for some asymmetries.

Equilibrium Outcomes and Strategies. Table 1 shows expected price levels, the standard deviation of prices and various welfare measures when we simulate data using equilibrium strategies in different periods of the finite horizon game. For comparison, expected joint-profit maximizing prices and static Nash equilibrium prices under CI (given average costs) are \$45.20 and \$22.62, with small standard deviations (\$0.007 and \$0.011). Signaling MPBE prices are higher and significantly more volatile than Nash prices when the game is more than a couple of periods from the end, but they are always much lower than joint profit-maximizing prices. We now describe the strategies that result in these outcomes.

Figure 1(a) shows four static BNE period T pricing functions for firm 2, for different values of firm 1's period $T-1$ marginal cost ($c_{1,T-1}$), assuming that both firms know/believe that $c_{2,T-1} = \$8$. Firm 2's price increases with $c_{1,T-1}$ as firm 1's expected period T price rises with $c_{1,T-1}$. However, the variation in firm 1's prior cost affects firm 2's price by less than one cent, and, averaging across all possible cost realizations, average prices and welfare are almost identical to outcomes with CI.[12] Therefore the existence of asymmetric information alone (i.e.,

[11]For example, the probability that a firm with the highest marginal cost has a cost in the lower half of the support in the next period is 0.32.

[12]Expected producer and consumer surplus differ by less than \$0.0001 across these models.

Figure 1: Period T and $T-1$ Pricing Strategies in the Finite Horizon, Continuous Type Signaling Game

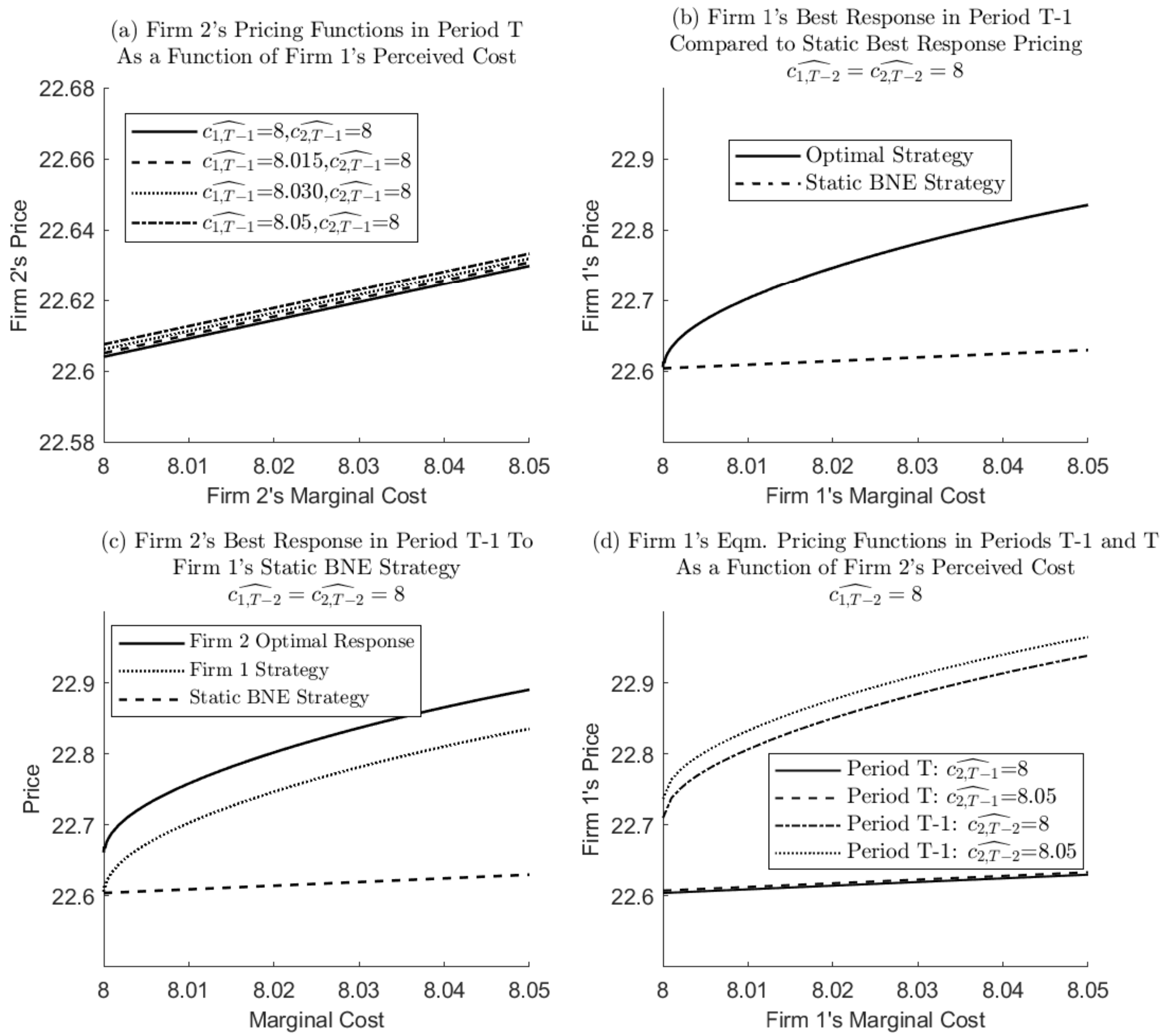


Table 1: Equilibrium Prices and Welfare in the Duopoly Game

				Expected Welfare Measures Per Market Size Unit		
Period	Nature of Equilibrium	Mean Price	Std. Dev. Price	Cons. Surplus	Producer Surplus	Total Welfare
T-24	MPBE	\$24.76	\$0.47	\$30.91	\$15.96	\$46.87
T-13	MPBE	\$24.76	\$0.47	\$30.91	\$15.96	\$46.87
T-10	MPBE	\$24.75	\$0.47	\$30.92	\$15.95	\$46.87
T-7	MPBE	\$24.68	\$0.45	\$30.98	\$15.89	\$46.88
T-4	MPBE	\$24.25	\$0.36	\$31.40	\$15.51	\$46.91
T-2	MPBE	\$23.38	\$0.17	\$32.23	\$14.74	\$46.97
T-1	MPBE	\$22.88	\$0.06	\$32.71	\$14.29	\$47.00
T	BNE	\$22.62	\$0.01	\$32.96	\$14.05	\$47.01
Infinite Horizon	Stationary MPBE	\$24.76	\$0.47	\$30.91	\$15.96	\$46.87

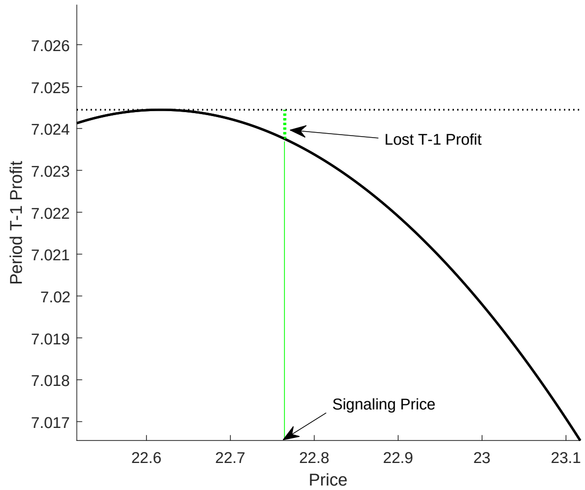
Notes: except for the last row, all prices are based on equilibrium strategies in a finite horizon model with parameters described in the text. The last line reports results for the stationary strategies in an infinite horizon model with the same parameters.

when not combined with some form of dynamics) does not generate interesting effects given our parameters.

There is an incentive to signal in period $T-1$ because a firm's price can affect its rival's price in period T. Assuming both firms' period $T-2$ costs were \$8, Figure 1(b) shows firm 1's signaling pricing function (found by solving the differential equation (1) given the boundary condition (2)) if it expected that firm 2 was using its period T strategy. We reproduce the period T pricing strategy for comparison. The pricing functions intersect for $c_{1,T-1} = \$8$, but signaling may lead firm 1 to raise its price by as much as 20 cents for higher costs. At first blush, this large increase may seem surprising given that we know the effect of any signal on firm 2's T price will be small. However, the assumed demand implies that firm 1's profit function, shown in Figure 2, is sufficiently flat that, if $c_{1,T-1} = \$8.025$, its expected lost period $T-1$ profit from using a signaling price of \$22.76, rather than the statically optimal period $T-1$ price of \$22.61, is only \$0.00070 per consumer, which is less than the (discounted) expected period T profit gain of \$0.00079 from being viewed as a firm with $c_{1,T-1} = \$8.025$ rather than $c_{1,T-1} = \$8.0001$, the cost firm 2 would infer if firm 1 set a price of \$22.61.

Figure 1(b) assumed that firm 2 was using its period T strategy with no signaling. Figure

Figure 2: Expected $T-1$ Period Profit Function: $c_{1,T-1} = \$8.025$ and $c_{1,T-2} = c_{2,T-2} = \$8$



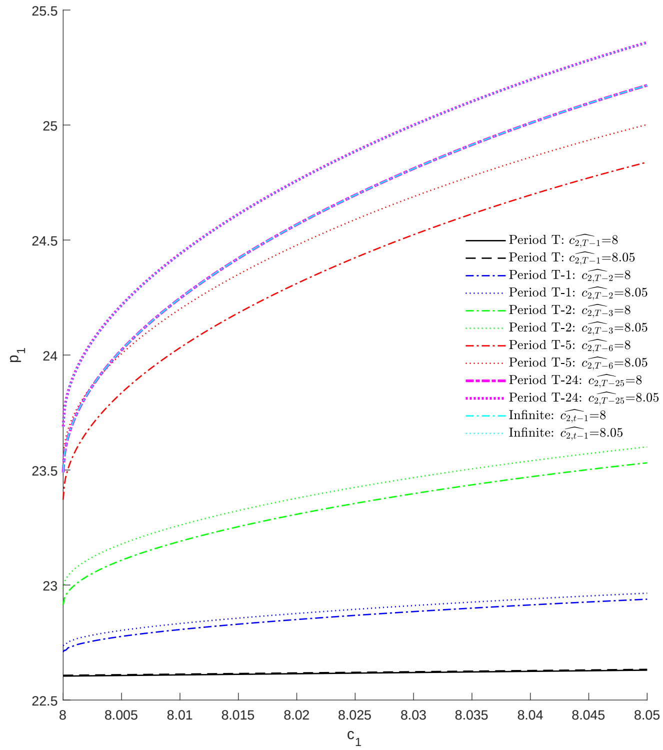
Notes: the profit function is drawn "per potential consumer" for a firm assumed to have a marginal cost of \$8.025, and with a rival using the static BNE pricing strategy when both firms' previous period marginal costs were \$8.

1(c) shows firm 2's best signaling response when firm 1 uses the strategy in Figure 1(b) (repeated in the new figure as a comparison). As firm 1's expected price has increased, firm 2's static best response pricing function shifts upwards. Of course, this positive feedback will cause firm 1's pricing function to rise as well. Figure 1(d) shows the equilibrium period $T-1$ pricing functions. The increase in the slope and the dispersion of the pricing functions means that period $T-1$ prices will be higher and more volatile than period T prices.

The increased vertical spread also means that period $T-1$ prices are more sensitive to perceived period $T-2$ costs which increases period $T-2$ signaling incentives. Figure 3 shows a selection of equilibrium pricing functions for period $T-2$ and earlier periods. The pricing functions become more spread out and the level of prices increases, although by successively smaller amounts, in earlier periods. Further back than period $T-15$ equilibrium pricing functions and average prices barely change. The figure also plots the stationary pricing strategies that we compute for an infinite horizon game with the same parameters. They are indistinguishable from the strategies in the early periods of the finite horizon game.[13]

[13]We have consistently found this convergence except in cases when the conditions required for separation are violated or are very close to being violated in which case the infinite horizon strategies may not converge.

Figure 3: Equilibrium Pricing Functions for Firm 1 in the Infinite Horizon Game and Various Periods of the Finite Horizon Game.



Notes: all functions drawn assuming that firm 1's perceived marginal cost in the previous period was \$8.

3.2 Cost Assumptions, Signaling Incentives and the Existence of Separating Equilibria.

As noted, signaling incentives in the previous example are relatively weak because of the limited correlation in marginal costs across periods. Increasing the AR(1) parameter or $\bar{c} - \underline{c}$, or reducing σ_c tend to increase signaling incentives and raise equilibrium prices. However, when price increases are too large, the conditions for characterizing best responses can fail and we may not be able to find a separating equilibrium.

The first six columns of Table 2 show, for different periods, the baseline average prices and average prices when signaling incentives are strengthened. Small parameter changes result in higher equilibrium prices, but larger changes result in the failure of our algorithm as we cannot define best response pricing functions. Pooling or partial pooling equilibria may exist, but we do not know how to characterize them. Appendix B.1.2 uses a two-type example to examine

Table 2: Equilibrium Pricing in a Finite Horizon Game with Alternative Cost Specifications

	Baseline	Expand Range			Reduce Std. Deviation		Expand Range & Increase Std. Dev.
$[\underline{c}, \overline{c}]$ (\$)	[8,8.05]	[8,8.075]	[8,8.15]	[8,8.3]	[8,8.05]	[8,8.05]	[8,8.50]
σ_c (\$)	0.025	0.025	0.025	0.025	0.02	0.01	0.25
T-24	\$24.76	\$26.51	-	-	\$25.71	-	\$24.90
T-10	\$24.75	\$26.59	-	-	\$25.70	-	\$24.89
T-9	\$24.74	\$26.59	fails	-	\$25.69	fails	\$24.89
T-8	\$24.72	\$26.57	\$28.48	-	\$25.66	\$28.58	\$24.89
T-7	\$24.68	\$26.50	\$29.17	fails	\$25.60	\$28.76	\$24.87
T-6	\$24.61	\$26.37	\$29.35	\$30.40	\$25.49	\$28.65	\$24.85
T-1	\$22.88	\$23.05	\$23.42	\$23.93	\$22.93	\$23.05	\$23.55
T	\$22.62	\$22.63	\$22.67	\$22.74	\$22.62	\$22.62	\$22.84
∞-Horizon	\$24.76	\$26.50	fails	fails	\$25.71	fails	\$24.90

Notes: values in all but the last line are based on the duopoly, continuous type, finite horizon model with demand parameters described in the text (cost parameters indicated in the table). The last line reports results for the stationary strategies in the infinite horizon model with the same parameters. "Fails" indicates that the belief monotonicity or single-crossing conditions fail so that we cannot calculate signaling best response pricing functions.

the failure of the conditions, including belief monotonicity, in more detail.

However, as illustrated in the final column, we can sustain separating equilibria if we increase $\overline{c} - \underline{c}$ and increase σ_c simultaneously.[14] This pattern will be relevant for our application, where we will estimate that both $\overline{c} - \underline{c}$ and σ_c are higher than in our baseline example.

3.3 Equilibrium Outcomes with Different Numbers of Firms, Asymmetries and Alternative Discount Factors.

A firm's strategic incentive to raise its current price in order to raise its rivals' future prices may be dulled when there are more firms because (i) price increases may be more costly as the firm's residual demand becomes more elastic, and (ii) an individual firm's signal will tend to have a smaller effect on the price that any other firm sets in the next period. We therefore consider how the magnitude of price differences between our model and a static CI model vary with the number of firms.

We use the two-type model so that we can consider up to seven, ex-ante symmetric single-product firms, but we stay as close as possible to our previous example by assuming the same

[14] In the final column the probability that a cost goes from one extreme of the support to the opposite half of the support is 0.32, which is the same as in the baseline case.

demand system and that each firm's marginal cost is either 8 or 8.05 with a probability of 0.3 of the marginal cost changing between periods. We consider 5 alternative values of the discount factor, $\{0.5, 0.8, 0.9, 0.95, 0.99\}$, to investigate whether effects are sensitive to assuming $\beta = 0.99$. As we identified some examples of multiple signaling equilibria in the infinite horizon two-type model, we present results from a finite horizon model where we solve backwards for additional periods until the strategies converge.[15]

Figure 4(a) shows average prices predicted by our signaling model and a CI model where the finite horizon assumption implies that firms will always set static Nash prices. Static pricing is always optimal when $N = 1$ so all specifications predict identical prices. When $N = 2$, average prices are 15.3% above CI Nash levels when $\beta = 0.99$ and 9.4% higher when $\beta = 0.8$, so it is possible to generate substantial effects even with a discount factor consistent with annual or even less frequent pricing. As N increases, signaling raises average prices by smaller, but not necessarily trivial, percentages: for example, when $N = 4$ and $\beta = 0.95$, signaling prices are 2% higher than CI prices.[16]

We also use a two-type example to examine whether symmetry is necessary to produce large effects using a three-firm specification. We continue to assume that each firm's marginal cost is either 8 or 8.05. We assume nested logit demand with nesting and price coefficients of 0.25 and -0.1, with firm-specific indirect utility intercepts calibrated so that, with average marginal costs for each firm and static CI Nash prices, the three firms have specific shares of sales and 97.5% of potential consumers make a purchase.[17] $\beta = 0.99$.

Figure 4(b) reports the increase in the share-weighted average price relative to CI Nash, with the circle areas indicating the magnitudes that are also written in the figure, where the share of the largest firm and the split of the shares of the other firms are represented on the axes. The price increase in the symmetric 3-firm model is 4.6% (bottom-left circle). The percentage increases are largest when the industry is close to an effective duopoly, but they are significant

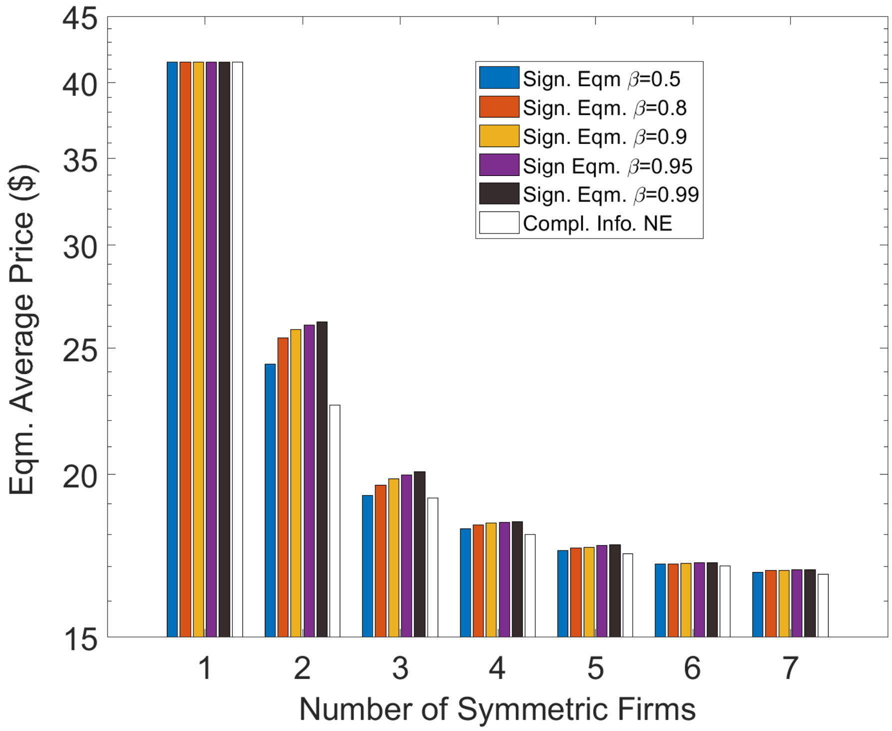
[15] Convergence is defined as a maximum difference in the pricing strategies across periods of less than 1e-4. In all cases we go back at least 30 periods, by which point the convergence criterion has been reached for most of our specifications.

[16] We have also calculated average joint-profit maximizing prices under CI, and the critical discount factor that would sustain collusion with Nash reversion trigger strategies if the game had an infinite horizon. For example, with $N = 4$, joint-profit maximizing prices of over 49 could be sustained if $\beta > 0.62$.

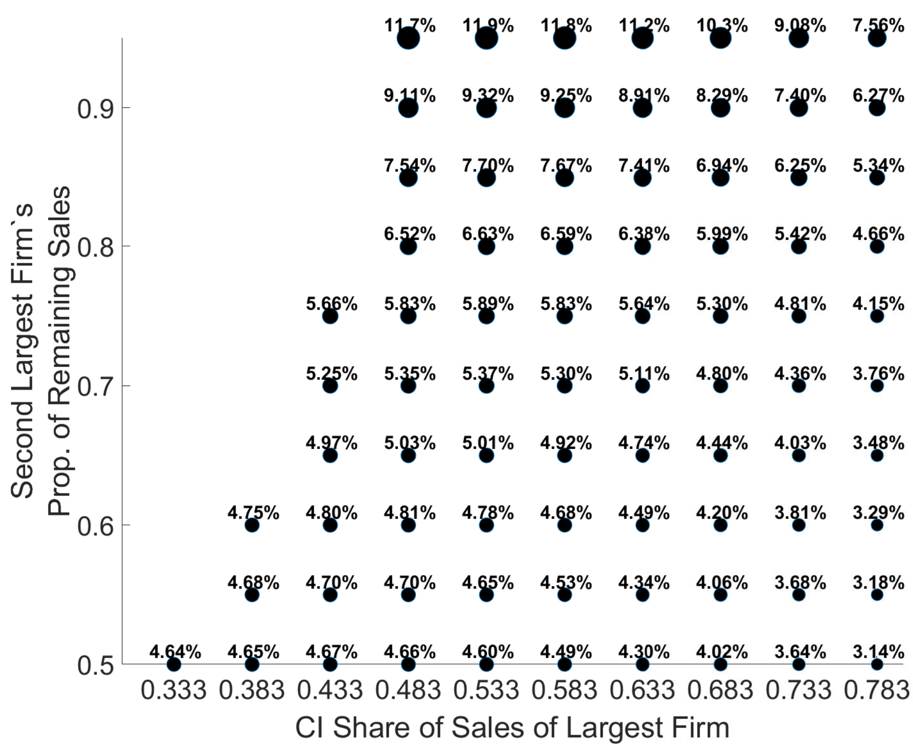
[17] The final assumption implies limited diversion to the outside good, where diversion is defined by the proportion of lost demand that goes to the outside good when the price of good rises. As our earlier discussion suggests, limited diversion to the outside good is necessary for us to find any separating equilibrium for the most extreme asymmetries that we consider.

Figure 4: Equilibrium Average Prices with Different Numbers of Firms and Alternative Discount Factors in a Two-Type Model, and the Effect of Demand-Side Asymmetries with Three Firms.

(a) Average Prices with Symmetric Single-Product Firms.



(b) Increase in Share-Weighted Average Prices Relative to CI Nash in a Three Firm Model with Demand Asymmetries.



in other cases as well. For example, when the CI shares of total sales are $\{0.68, 0.24, 0.08\}$, the average price is 5.3% higher than the average CI Nash price.[18]

3.4 Mergers and Merger Analysis.

Signaling may matter in any application where a small number of oligopolists set prices, but we focus on the implications for horizontal mergers because static CI Nash assumptions underpin almost all of the quantitative predictions that agencies conduct.

Figure 5(a) uses the same specifications as Figure 4(a) but reports the increase in average firm prices after an unanticipated merger of two firms which eliminates a product without generating synergies. As there is no signaling in monopoly, a 2-to-1 merger increases prices more under CI Nash and with lower discount factors. However, price increases are larger with signaling than with CI Nash when the discount factor is large enough for the remaining mergers.[19]

It is more common to assume that a merged firm will continue to sell both products after a merger, and that synergies are possible. In order to maintain the tractable single unobserved state variable-single signal structure of our model, we will assume that, post-merger, the merged firm will have exactly the same marginal cost for both products in each period and it will have to set a single price. The cost evolution process will be unaffected, even if the level of marginal cost falls for the merging firm. To make this assumption plausible, we will only consider mergers involving products with the same indirect utility intercepts.

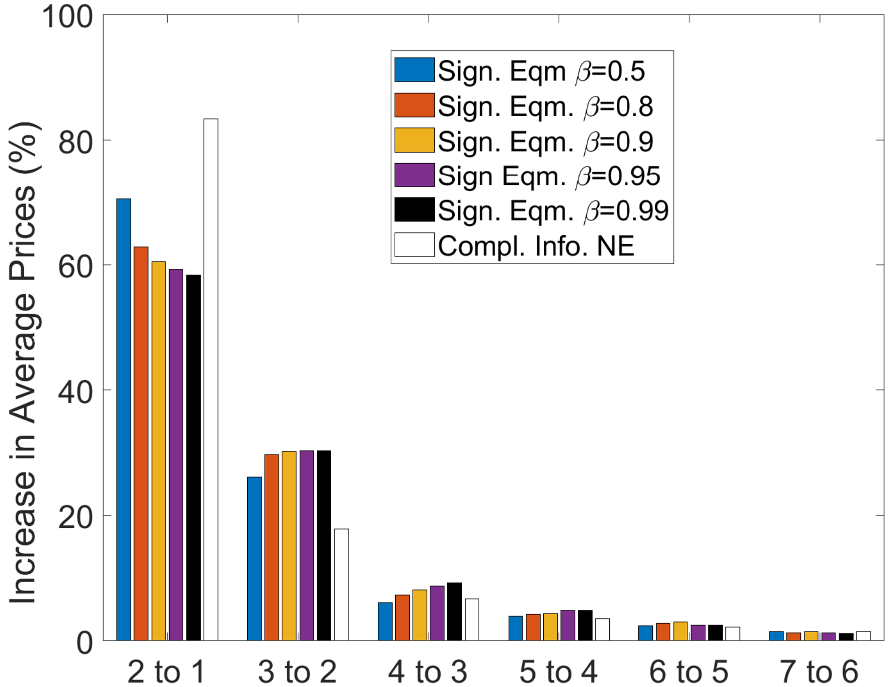
Figure 5(b) considers the effect of mergers in asymmetric 4-firm industries. The pre-merger model extends the example used for Figure 4(b) to an additional firm. The x-axis indicates the combined market share of the two merging firms (under CI Nash) before the merger (so 0.5 means that each firm makes 25% of sales), while the y-axis shows the split of the remaining CI sales between the two remaining firms. The price increases shown in the figure are the average price increase (i.e., the increase in the share-weighted average price across the four products) when the merged parties benefit from a marginal cost synergy which would, at average marginal

[18]In this example, it is the firm with the second largest share that increases its average price the most, in both dollar and percentage terms (10.0%).

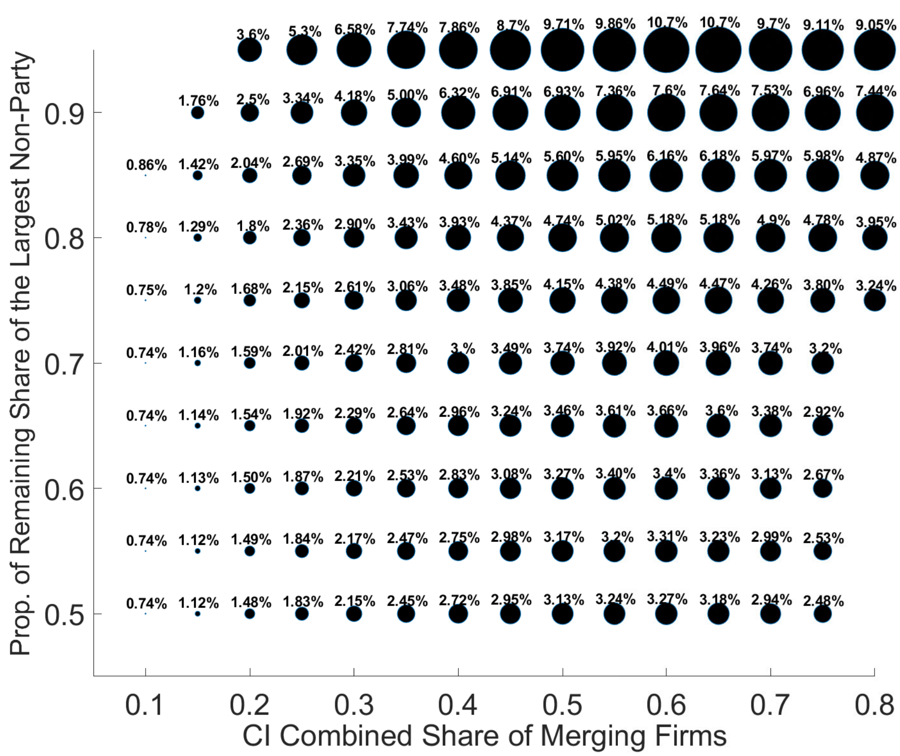
[19]In these examples, the absolute price increase is always larger with signaling, but the proportional price increasea can be smaller because pre-merger signaling prices are higher. This fact also explains why price increases in 6-to-5 and 7-to-6 mergers are not monotonic in the discount factor.

Figure 5: Effects of an Unanticipated Merger in a Two-Type Model.

(a) Increases in Average Prices with Symmetric Single-Product Firms Before and After the Merger.



(b) Increases in Average Prices in a Four Firm Model where Merged Firm Has Two Products after the Merger and Benefits from the CI CMCR. Missing values indicate that a separating equilibrium was not found before or after the merger.



costs, prevent any post-merger price increase if firms played CI Nash. This level of synergy is typically known as the Compensating Marginal Cost Reduction (CMCR). Any price increase would therefore be a surprise to an agency that, based on a CI model, would expect the merger to be competitively neutral.

A merger in a symmetric four-firm industry generates a surprise price increase of 3.13%. Configurations where the rest of the market is dominated by a single firm can produce much larger price increases. This reflects important differences in economics of mergers with CI and with signaling. As shown in recent work by Nocke and Whinston (2020), CMCRs will typically depend on the market shares of the merging firms, but not the concentration of the rest of the market.[20] In contrast, with signaling, non-merging rivals, especially large ones, have incentives to raise prices when, after a merger, the prices of the merged firm will be more sensitive to the prices that the rivals set (even if, hypothetically, the average level of the merged firm's price was to be unchanged). Of course, the merged firm may respond to rivals' price increases with price increases of its own, creating a positive feedback which leads to substantial price effects.

For example, if pre-merger sale shares are $\{0.325,0.325,0.33, 0.02\}$ the merged firm, which benefits from the synergy, raises its average price by 10.6%, and the large rival increases its average price by 12.9% (the small rival's average price increases by 1%). A static CI Nash model where the static best response functions have slopes less than one (and usually slopes are significantly less than one) can never predict that a rival will increase its prices by more than the merging firm whatever level of synergy is assumed. On the other hand, this can often happen in our model.

[20]The result is exact for multinomial logit demand, but Nocke and Whinston show that the result holds approximately in the random coefficients nested logit model considered by MW in their analysis of beer, which is, of course, closely related to our empirical application.

Table 3: Post-Merger Prices and Required Synergies in an Infinite Horizon Continuous-Type Model. Firms are symmetric before the merger, and the merged firm sells two products after the merger.

	4-to-3 Merger	3-to-2 Merger
Pre-Merger Average Prices	\$18.25	\$19.79
Post-Merger Average Price of Merged Firm if No Marginal Cost Synergy	\$21.53 (+18.0%)	\$27.18 (+37.3%)
Post-Merger Average Price of Non-Merging Firm if No Marginal Cost Synergy	\$19.12 (+4.8%)	\$23.59 (+19.2%)
CI CMCR	\$4.95	\$10.08
Merged's Firm Post-Merger Average Price with CI CMCR Synergy in Signaling Equilibrium	\$18.85 (+3.2%)	\$23.00 (+16.2%)
Marginal Cost Reduction Required to Keep Merged Firm's Average Price from Rising in Signaling Equilibrium	\$5.76	\$19.94

Notes: parameterization described in the text. Note that the CI CMCR is the marginal cost reduction that an analyst would compute using the true demand system, observed (signaling) pre-merger signaling prices and a CI Nash assumption.

These effects can mean that the marginal cost reduction that would actually be required to prevent a price increase may need to be substantially higher than the CI CMCR, and this is true even when firms are symmetric before the merger. We illustrate this using a final example where we extend our infinite horizon continuous-type model to allow up to four firms, with the same assumptions on demand and marginal costs. Firms are symmetric before the merger, but after the merger the merged firm has two products. Table 3 shows the price effects of 4-to-3 and 3-to-2 mergers. With no synergy, either merger leads to both the merging firms and the non-merging firms increasing their prices substantially. To prevent the merged firm's price from rising above its average pre-merger level, very large synergies are required. In fact for a 3-to-2 merger, the merged firm's average marginal cost would have to fall from \$8.025 to -\$11.915.

3.5 Additional Examples.

Appendices B.1.1 and B.2 describe additional examples. Appendix B.1.1 uses the two-type duopoly model to examine the relationship between the existence of separating equilibria, the effects of signaling on prices, the serial correlation of costs and the extent to which, when a firm's price rises, demand is diverted to the outside good. Price increases above static CI Nash can be very large (an increase of 45% in one case) when there is limited diversion to the outside good even when there is moderate serial correlation in costs (e.g., $\Pr(c_{i,t} = c_{i,t-1})$=0.75). On the other hand, when there is significant diversion to the outside good, separating equilibria can only be supported when there is lower serial correlation in costs, and equilibrium price increases are smaller.

Appendix B.2 present three simple duopoly examples where marginal costs are fixed and known, but firms have private information about some other element of their payoff function (a feature of demand, the weight managers place on revenues rather than profits, or the weight they place on the profits of rivals). Signaling can raise prices significantly above CI Nash levels in each case. Therefore, while our empirical application will assume that it is marginal costs that are privately known, one would still get substantial price effects if it is a different part of firms' profit functions that are opaque to rivals.

4 Empirical Application: The MillerCoors Joint Venture

In this section, we test whether our model can explain changes in price levels, and price dynamics in the U.S. beer industry around the time of the 2008 MC JV, which was an effective merger of the second- and third-largest US brewers.[21] We structure our discussion by first discussing our motivation for looking at this setting, before describing the calibration of our model, and a CI alternative, using pre-JV data. We compare what these models predict should have happened to prices after the JV with what we observe in the data. Discussion of the data, which is the same as used by MW, demand estimation and additional analysis of collusive models that assume CI are contained in Appendix D in order to keep the text appropriately focused on our model.

[21] Anheuser-Busch was purchased by InBev in 2008. Throughout this section and the Appendices, we will use AB to refer to Anheuser-Busch before 2008 and Anheuser-Busch InBev afterwards, and we will assume that this transaction had no effect on AB's pricing incentives.

4.1 Motivation.

There are three broad reasons why we choose to consider the MC JV even though our model could be applied to other transactions affecting differentiated product markets.[22]

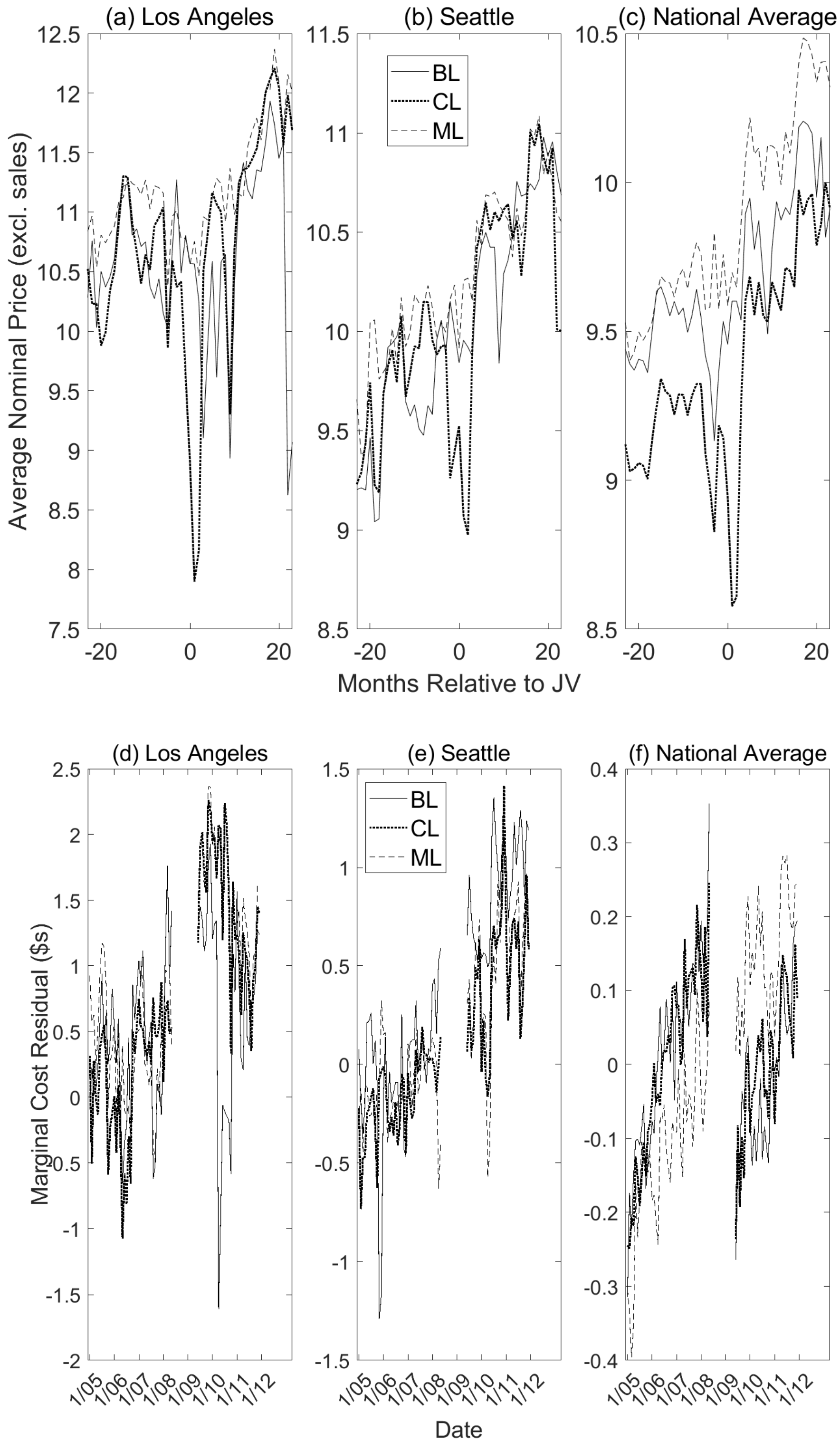
First, features of the transaction and the industry mean that the assumptions that we make for tractability are relatively plausible. Immediately prior to the JV, the leading brands of Miller and Coors had very similar market shares, at least at a national level, and sold at very similar retail prices (Appendix D.3). The prices of different major brands sold by the same brewer (e.g., Budweiser and Bud Light) tend to move together (Appendix D.6). These facts make us comfortable with assuming that Miller and Coors are symmetric before the JV and that each pre-JV firm sets a single price. After the JV, MC produced Miller and Coors brands, including Miller Lite (ML) and Coors Lite (CL), in the same breweries, so that their production and distribution costs should be almost identical and move together, and in the data ML and CL prices become more correlated (Figure 6(a)-(c) below, and Appendix D.6). These facts are consistent with assuming that MC has a single marginal cost and a single price after the JV. The "subpremium" and "premium" segments of the beer industry were also dominated by three domestic firms (two after the JV) whereas import and domestic craft beers are sold at significantly higher prices (Appendix D.3). There is no evidence that post-JV retail price increases for the domestic brands caused significant substitution to these higher-priced alternatives (Appendices D.3 and D.5). We therefore believe that a model that only has the three large domestic brewers provides a reasonable, if imperfect, approximation of the market before the JV.

Second, a review of price data suggests that there are two features that we would like a model to explain. The first feature is the increase in the prices of domestic brands, including those produced by non-party AB, after the JV despite MC realizing marginal cost synergies (Ashenfelter, Hosken, and Weinberg (2015)). Regressions in Appendix D.4 quantify these price increases to lie between 40 cents and a dollar per 12-pack, or 3%-6%, depending on the specification. We will proceed assuming that MW's interpretation that the relative price increase was a causal anticompetitive effect of the JV is correct.[23] As discussed in Section 3.4,

[22] In ongoing work we are also applying the model to analysis of firms that choose capacities in the presence of asymmetric information.

[23] This interpretation is complicated by how the Great Recession may have affected demand and a fall in the

Figure 6: Average Prices and Marginal Cost Residuals for Flagship Brand 12-Packs in Three Markets. Panels (a)-(c) show nominal weekly average prices for 48 months around the JV, excluding sales at temporary price reductions. See Appendix D.7 for alternative versions. Panels (d)-(f) show the monthly marginal cost residuals implied by MW's RCNL-1 demand and supply model. MW exclude the 12 months following consummation of the merger.



our model can predict that a large rival will increase prices as much as, or more than, the merging parties.

The second feature is that average retail prices are quite volatile both before and after the JV. This is illustrated in Figure 6(a)-(c) which show monthly average nominal retail prices (calculated as dollar sales divided by units sold) for Bud Light (BL), ML and CL 12-packs in two local markets and nationally.[24] We exclude sales at prices that the IRI data indicate as temporary price reductions from our calculations, as these may create volatility that does not reflect changes in wholesale prices, although it is also possible that some price reductions and other types of promotion are funded by brewers or their distributors. Appendix D.7 shows that volatility remains when prices are calculated in different ways, including as the unweighted average of prices across stores so that shifts in volumes across sample stores do not affect the averages. Tacit collusion models either say nothing about price volatility, or when they are extended to allow for asymmetries of information, they may predict that prices should be rigid (Athey, Bagwell, and Sanchirico (2004)). On the other hand, relatively small variations in an underlying firm state variable can lead to significant price volatility in our model. Figure 6(d)-(f) plots the marginal cost residuals, for the same products and markets, implied by MW's "RCNL-1" demand and supply estimates where the supply model assumes CI and includes a conduct parameter after the JV. The implied marginal costs are volatile and serially correlated, which will be a feature that will be embedded into our model.[25]

Third, existing explanations for why prices increased after the JV based on CI theories of firm conduct can be shown to be inconsistent with the data. MW assume static CI Nash pricing before the JV and estimate a post-JV conduct parameter, which allows for partial joint profit-maximization by the domestic firms. MSW propose a more formal model of CI tacit collusion where, each year in each local market, a domestic price leader suggests an incentive-compatible supermarkup that the domestic firms should add to CI Nash prices. They also estimate a parameter, that is intended to capture factors such as the time preferences of the

CPI-U deflator, from 220.0 in July 2008 to 210.2 in December 2008, at exactly the same time that the merger was being consummated.

[24]We present nominal prices so that the picture is not distorted by the drop in the CPI deflator (footnote 23). When one examines nominal prices, volatility is arguably a more striking feature of the price series than the post-JV price increase.

[25]The ML and CL residuals become very similar in Los Angeles and Seattle after the JV, which provides additional support for our post-JV modeling assumption.

domestic firms, that rationalizes the observed price increases. While MW's and MSW's extra parameters are exactly identified by how AB increases its prices, assuming that the JV did not affect AB's marginal costs, we test MW's and MSW's assumptions using additional exclusion restrictions and including additional controls in Appendix D.9. We reject MW's assumption of CI Nash pricing before the JV in many specifications, as well as MSW's assumption of a common domestic supermarkup, while not necessarily rejecting the hypothesis that domestic brewer conduct was the same before and after the JV. While these results do not imply that no CI model can explain the data, it does suggest that alternative theories, including ours, deserve serious consideration.[26]

Our empirical approach is different to the one used by MW or MSW. They identify conduct or time preference parameters that rationalize the observed post-JV increase in prices, and do not seek to empirically test or otherwise validate their assumption that domestic brewers collude.[27] In contrast, we calibrate our model using **only** pre-JV data and then examine what our model predicts about the effects of the JV on price levels and price dynamics without estimating additional parameters.

4.2 Calibration of the Dynamic Asymmetric Information Model.

We calibrate an infinite horizon, continuous marginal cost three-firm/product version of our model using pre-JV data, and then compare its predictions with post-JV data. For comparison purposes, we also calibrate a CI model that assumes that firms use static Nash pricing strategies using the same parameters and moments. We say "calibration", even though we estimate five cost parameters, because of the strong assumptions we make to limit the computational burden. The most important simplification is that our calibration will treat data from different markets as data from independent repetitions of the same game, rather than reflecting markets with different demand and cost primitives.

[26]One can interpret folk theorems as implying that some tacit collusion model that assumes CI is likely to match the data.

[27]MW and MSW refer to excerpts from AB's documents published by the DOJ in various complaints that can be read as suggesting collusive behavior. For example, the DOJ's Complaint enjoining AB's acquisition of Grupo Modelo highlighted that AB had a Conduct Plan for setting prices which aimed to "dictate consistent and transparent competitive response" and to yield the "highest level of followership". However, these quotes are not inconsistent with domestic brewers using prices to signal information about either their costs or their future competitive intentions, and they do not show that Nash reversion, "tit-for-tat", or one of the other types of punishment that tacit collusive models assume, are being threatened.

4.2.1 Products.

We model the pricing of three brands. We label these brands as BL, ML and CL, and will estimate the cost parameters to match the observed price dynamics of these flagship products. However, Appendix D.6 shows that the prices of brands in the same portfolio (e.g., Budweiser and BL) are highly correlated, and one can also view the brands as representing the portfolios of AB, Miller and Coors. Products of other brewers, including imports and craft beers, are included in the outside good.[28] We will assume that ML and CL are symmetric before the JV, as we will have to assume that MC sets the same price for both products after the JV.

4.2.2 Demand.

We assume static, time-invariant nested logit demand, with the three brands in the same nest. The parameters are the nesting and price parameters, and the mean utilities (excluding the effect of price) of BL and ML/CL. Our baseline parameters are chosen so that, at average real prices in the pre-JV data, the average own price elasticity is -3, the market shares of the three products are 28% for BL and 14% each for ML/CL and, on average, if the price of one brand increased, 85% of the demand that it loses would go to the other brands with the remainder to the outside good.[29] See Appendix D.8 for empirical estimates of demand that are consistent with these assumptions. When we use weekly data on 6/12/18/24/30-packs and exclude temporary price reductions, the pre-JV cross-market average prices are \$10.09 for BL and \$9.95 for ML/CL, and the implied nesting and price parameters are 0.772 and -0.098, and the BL and ML/CL mean utilities are 1.044 and 0.863 respectively.

4.2.3 Marginal Costs.

We assume that the marginal costs of product i, $c_{i,t}$, lie on the interval $[\underline{c_i}, \underline{c_i} + c']$, where we estimate $\underline{c_{BL}}$, $\underline{c_{ML/CL}}$ and c'. $c_{i,t}$ evolves according to an AR(1) process with truncated innovations

$$c_{i,t} = \rho c_{i,t-1} + (1-\rho)\frac{\underline{c_i} + \underline{c_i} + c'}{2} + \eta_{i,t} \tag{4}$$

[28] An earlier version estimated a model that included imports as a non-signaling fringe. The calibrated model predicted that, after the JV, they would raise their prices by around 2 cents.

[29] These assumed shares overstate the share of BL relative to ML and CL, but understate the share of AB, relative to Miller and Coors, in the beer market and the light beer segment.

where $\eta_{i,t} \sim TRN(0, \sigma_c^2, \underline{c_i} - \rho c_{i,t-1} - (1-\rho)\frac{\underline{c_i}+\underline{c_i}+c'}{2}, \underline{c_i} + c' - \rho c_{i,t-1} - (1-\rho)\frac{\underline{c_i}+\underline{c_i}+c'}{2})$ and σ_c is the standard deviation of the untruncated innovation distribution. The fit of the model improves only slightly if we allow ρ, σ_c and c' to vary across firms.

4.2.4 Objective Function, Matched Statistics and Identification.

The cost parameters are estimated using indirect inference (Smith (2008)). For a given value of the cost parameters, we solve the model (see Appendix A.2 for the method) and simulate a time-series of data to calculate six statistics/regression coefficients that we match to ones from the data that we describe below. The estimation problem is

$$\widehat{\theta} = \arg\min_{\theta} g(\theta)' W g(\theta)$$

where $g(\theta)$ is a vector where each element k has the form $g_k = \frac{1}{M} \sum_m \tau_{k,m}^{data} - \widehat{\tau_k(\theta)}$ where $\tau_{k,m}^{data}$ is a statistic estimated using the actual data and $\widehat{\tau_k(\theta)}$ is the equivalent statistic estimated using simulated data from the model solved using parameters θ. W is a weighting matrix. The reported results use an identity weighting matrix, although the choice of W has little effect on the parameters as we match all of the statistics almost exactly. The objective function is minimized using `fminsearch` in MATLAB (version 2018a). Standard errors are calculated treating different markets before the JV as independent observations on the same game. Estimation takes between 12 and 24 hours.[30]

For each geographic market, we calculate six statistics using data from January 2001 to the announcement of the JV in October 2007.[31] Our preferred specification uses weekly data and the five most common pack sizes (6, 12, 18, 24 and 30-packs).[32] Market-week-brand-size average real prices per 12-pack equivalent are calculated excluding temporary store price reductions, and using only market-weeks where we observe more than five stores.[33] The first two statistics that we match are the (unweighted) average prices for BL and ML across pack sizes and weeks.

[30]Computationally light two-step approaches, which are often used to estimate dynamic games, cannot be used because they require that all serially-correlated state variables, which in our setting would include beliefs, are observed by the researcher.

[31]MW use data from 2005 when estimating demand because they need to match sales data to data on market demographics. We use a longer sample as we do not include demographics in our estimation of the supply-side.

[32]Our model does not have different pack sizes, market heterogeneity, varying sets of stores or time trends, so the regressions using simulated data do not control for these factors.

[33]See Appendix D for a discussion of the sample selection.

The third statistic is the interquartile range (IQR) of prices for BL. This is calculated as the IQR of the residuals for each market from a regression where, pooling markets, we regress the week-market-size prices of BL products on dummies for the specific set of stores observed in the market-week (interacted with pack size) and week-size fixed effects in order to control for fixed retail price differences across stores and any national promotions. The remaining statistics are coefficients from market-brand-specific regressions of market-week-brand-size prices on the lagged prices of all three brands. Specifically we use the averages of $\rho^{ML,ML}$ and $\rho^{CL,CL}$, $\rho^{BL,CL}$ and $\rho^{BL,ML}$, and $\rho^{ML,CL}$ and $\rho^{CL,ML}$, where $\rho^{i,j}$ is the coefficient on the lagged price of brand j when the dependent variable is the price of brand i. These AR(1) regressions include dummies for the exact set of stores observed, interacted with pack size, and a linear time trend.

Assuming that the equilibrium is unique, the intuition for identification is straightforward.[34] Given the assumed demand parameters and the observed price levels, the mark-ups implied by the model will identify the lower bounds on brand marginal costs. The AR(1) coefficients and the dispersion of prices will identify the range of costs and the parameters of the cost innovation process.[35] We will compare additional statistics that we do not match during estimation to understand the fit of the model.

To provide a sense of the AR(1) coefficients, Table 4 shows the coefficients from similar regressions that pool data from all markets for four alternative samples. Panel (a) reports the results for our preferred specification. The serial correlation parameters for a product's own price are between 0.41 and 0.46, while the cross-product correlations are positive but smaller. If price reductions are included (panel (c)), serial correlations fall, which is consistent with sales lasting one week and being proceeded and followed by higher regular prices. Serial correlation is higher if we use only 12-packs (panel (b)). Panel (d) repeats (a) using monthly prices and

[34]The possibility that our game has multiple equilibria may create two issues for estimation. First, the objective function may be hard to minimize if our solution algorithm jumps between different sections of the equilibrium correspondence. In practice, we can match our moments almost exactly across many alternative parameterizations. Second, another equilibrium supported by different parameters might give similar predictions to the equilibrium that our algorithm finds. This is essentially a potential identification problem. Here we have to rely on the fact that we have never found multiple equilibria in continuous-type games, although we suspect that they may exist for some parameters.

[35]Larger cross-brand ρ coefficients imply stronger signaling effects, so that a smaller range of costs may be required to generate the dispersion of prices in the data. Our experience is that we need to match average prices, some measures of own-brand and cross-brand serial correlation, and some measure of either the dispersion of prices or the variance of innovations in prices to identify the parameters. The exact combination of moments used has little effect on the results.

Table 4: Pre-JV AR(1) Price Regressions Using Flagship Market-Pack Size-Week or -Month Data

(a) Week, Price Reductions Excluded, All Pack Sizes, Fixed Effects for Set of Stores

	(1) $p_{BL,t}$	(2) $p_{ML,t}$	(3) $p_{CL,t}$
$p_{BL,t-1}$	0.451	0.056	0.043
	(0.033)	(0.017)	(0.010)
$p_{ML,t-1}$	0.030	0.409	0.016
	(0.011)	(0.036)	(0.014)
$p_{CL,t-1}$	0.027	0.021	0.461
	(0.012)	(0.015)	(0.040)
Observations	36,659	36,670	36,700
R-squared	0.979	0.972	0.978
Mean Price ($)	10.08	9.95	9.94
SD residuals ($)	0.184	0.221	0.197

(b) Week, Price Reductions Excluded, 12 Packs Only, Fixed Effects for Set of Stores

	(1) $p_{BL,t}$	(2) $p_{ML,t}$	(3) $p_{CL,t}$
$p_{BL,t-1}$	0.489	0.071	0.028
	(0.032)	(0.026)	(0.018)
$p_{ML,t-1}$	0.062	0.505	0.028
	(0.013)	(0.038)	(0.012)
$p_{CL,t-1}$	0.004	0.016	0.549
	(0.012)	(0.015)	(0.043)
Observations	10,829	10,817	10,828
R-squared	0.964	0.945	0.957
Mean Price ($)	10.30	10.22	10.19
SD residuals ($)	0.144	0.183	0.163

(c) Week, Price Reductions Included, All Pack Sizes, Fixed Effects for Set of Stores

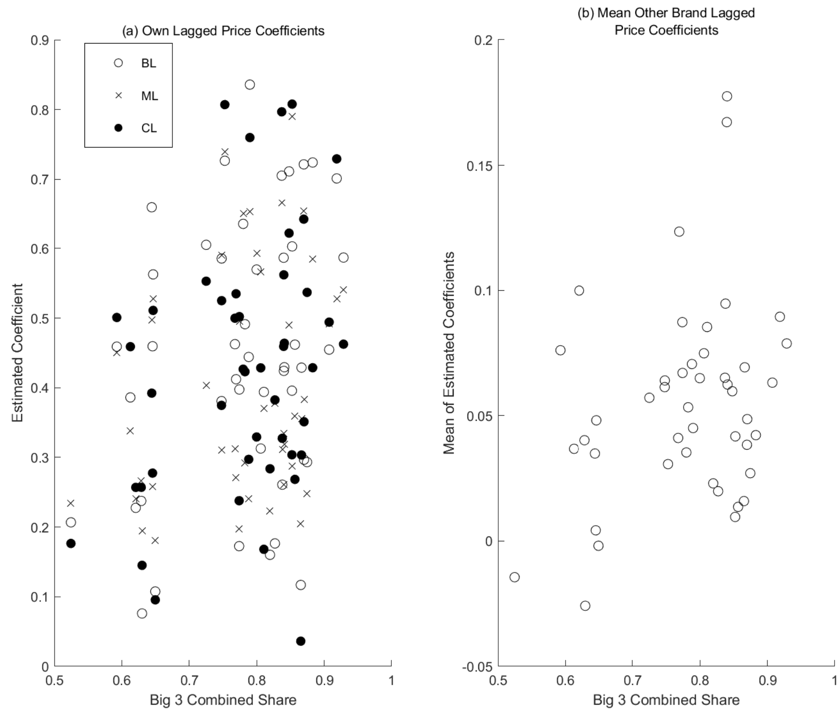
	(1) $p_{BL,t}$	(2) $p_{ML,t}$	(3) $p_{CL,t}$
$p_{BL,t-1}$	0.287	0.036	0.020
	(0.027)	(0.013)	(0.013)
$p_{ML,t-1}$	0.045	0.322	0.010
	(0.009)	(0.027)	(0.012)
$p_{CL,t-1}$	-0.023	-0.049	0.267
	(0.013)	(0.020)	(0.039)
Observations	37,449	37,431	37,442
R-squared	0.939	0.941	0.942
Mean Price	9.79	9.67	9.68
SD residuals	0.337	0.342	0.336

(d) Month, Price Reductions Excluded, All Pack Sizes, Fixed Effects for Markets

	(1) $p_{BL,t}$	(2) $p_{ML,t}$	(3) $p_{CL,t}$
$p_{BL,t-1}$	0.646	0.097	0.091
	(0.025)	(0.015)	(0.012)
$p_{ML,t-1}$	0.074	0.601	0.066
	(0.015)	(0.027)	(0.014)
$p_{CL,t-1}$	0.100	0.097	0.682
	(0.010)	(0.016)	(0.025)
Observations	13,972	13,973	13,975
R-squared	0.974	0.971	0.974
Mean Price	10.08	9.95	9.94
SD residuals	0.210	0.229	0.216

Notes: regressions also include time period*pack size interactions and use pack sizes containing volumes equivalent to 6, 12, 18, 24 and 30 12 oz. containers. Data from January 2001 to the announcement of the JV. Market or store fixed effects described in the label to each panel. Standard errors, clustered on the market, are in parentheses. The SD residuals statistic is the standard deviation of the residuals from the regression.

Figure 7: Estimated Pre-JV Price Dynamics and the Combined Market Shares of AB, Miller and Coors.



Notes: The estimated univariate regression coefficients, with standard errors in parentheses, for panel (a) are BL: 0.011 (0.226) + $0.558C_3$ (0.288), R^2 = 0.080; ML : 0.044 (0.192) + $0.465C_3$ (0.245), R^2=0.077; CL : -0.025 (0.215) + $0.568C_3$ (0.278), R^2=0.091; and for panel (b): -0.039 (0.046) + $0.120C_3$ (0.058), R^2=0.088.

market, rather than group-of-store, fixed effects (equivalent regressions will be used in our monthly data specification). In this case, the serial correlation parameters increase, but further investigation reveals that this happens primarily due to the change in the fixed effects.[36]

While our calibration does not seek to match cross-market heterogeneity, we note that the serial correlation coefficients show a pattern across markets that is consistent with our model. Using data simulated from our model, we typically estimate higher serial correlation parameters in price regressions when we change the parameters to induce larger signaling effects on prices, by, for example, reducing diversion to the outside good. Given any type of logit or nested logit preferences, diversion to other brands will tend to be lower when, as a group, the signaling brands have a higher market share. Figure 7(a) shows scatter plots of the estimated

[36]We have estimated monthly regressions including set of store fixed effects and dropping market-months where the set of stores changes within months. This causes the number of observations to drop dramatically: for example, the number of observations in the BL regression falls to 2,806, and the estimated coefficient on p^{BL}_{t-1} falls to 0.318. For some individual markets, there is not enough data to estimate serial correlation coefficients.

Table 5: Parameter Estimates for Seven Specifications.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Model	Signal	Signal	Signal	Signal	Signal	Signal	CI
Data Frequency	Week	Week	Week	Week	Week	Month	Week
Sizes	All	12 only	All	All	All	All	All
Price Reductions	Excl.	Excl.	Incl.	Excl.	Excl.	Excl.	Excl.
Mean Brand Price Elasticity	-3	-3	-3	-2.5	-3.5	-3	-3
Mean Flagship Diversion	85%	85%	85%	90%	80%	85%	85%
Lower Bound Cost for BL	\$5.259	\$5.278	\$4.845	\$4.248	\$5.973	\$4.616	\$5.439
($\underline{c_{BL}}$)	(0.222)	(0.048)	(0.046)	(0.043)	(0.026)	(0.127)	(0.010)
L.B. Cost for ML/CL	\$6.426	\$6.528	\$5.984	\$5.786	\$6.874	\$5.711	\$6.631
($\underline{c_{ML/CL}}$)	(0.094)	(0.014)	(0.022)	(0.024)	(0.017)	(0.020)	(0.058)
Width Cost Interval	\$0.600	\$0.752	\$1.246	\$0.556	\$0.672	\$1.793	\$0.672
($\overline{c_i}-\underline{c_i}$)	(0.043)	(0.021)	(0.018)	(0.102)	(0.026)	(0.037)	(0.097)
Cost AR(1) Parameter	1.178	0.939	0.850	1.222	0.959	0.742	1.088
(ρ)	(0.028)	(0.011)	(0.026)	(0.013)	(0.012)	(0.025)	(0.038)
SD Cost Innovations	\$0.262	\$0.278	\$0.566	\$0.260	\$0.270	\$0.400	\$0.278
(σ_c)	(0.031)	(0.001)	(0.050)	(0.104)	(0.026)	(0.052)	(0.086)

Notes: BL = Bud Light, ML = Miller Lite and CL=Coors Light. Standard errors in parentheses. The data specifications using weekly data include group-of-store fixed effects when calculating the data statistics. For the monthly specification, the regression using the data only include market fixed effects. Flagship diversion refers to the proportion of lost demand that switches to the other two products when the price of one of the product increases.

market-level serial correlation parameters for BL, ML and CL against the share of all beer sales accounted for AB, Miller and Coors in 2007. Figure 7(b) shows a similar plot for the average of the six cross-brand coefficients. In both cases there is a positive, and, using a regression analysis, a statistically significant, relationship, consistent with our simulations.[37]

We also calibrate the cost-side parameters assuming that firms have CI (i.e., they know each other's marginal costs) and use static Nash pricing strategies. This is done using the same procedure and moments that we use for the signaling model.

4.2.5 Parameter Estimates and Model Fit.

Table 5 reports the calibrated parameters for six signaling models, where different demand parameters are assumed or different prices series are matched, and one CI specification. For the signaling specifications, estimated marginal costs increase when demand is more elastic, and the range of costs and the standard deviation of the innovations increase when we match

[37]We also find positive, statistically significant relationships when we look at individual cross-brand coefficients.

Table 6: Model Fit for Three Specifications Using Weekly Data, Average Brand Price Elasticity of -3 and Flagship Diversion of 85%

Data Freq.		Week			Week
Sizes		All			12
Price Reductions		Excl.			Excl.
	Data	Sign. Model	CI Model	Data	Sign. Model
			Matched Moments		
Mean p_{BL}	\$10.09	\$10.09	\$10.09	\$10.30	\$10.30
Mean p_{ML}	\$9.96	\$9.96	\$9.96	\$10.22	\$10.22
Mean $\rho^{ML,ML}$, $\rho^{CL,CL}$	0.402,0.413	0.408	0.407	0.468,0.450	0.444
Mean $\rho^{BL,ML}$, $\rho^{BL,CL}$	0.082,0.066	0.074	-0.000	0.102,0.056	0.076
Mean $\rho^{ML,CL}$, $\rho^{CL,ML}$	0.051,0.036	0.046	0.005	0.065,0.026	0.035
IQR p_{BL}	\$0.189	\$0.189	\$0.189	\$0.185	\$0.212
			Unmatched Moments		
Mean p_{CL}	\$9.95	\$9.97	\$9.97	\$10.20	\$10.23
$\rho^{BL,BL}$	0.444	0.403	0.412	0.442	0.418
Mean $\rho^{ML,BL}$, $\rho^{CL,BL}$	0.059,0.0.42	0.038	-0.002	0.065,0.040	0.038
SD of BL Res.	\$0.177	\$0.107	\$0.111	\$0.136	\$0.122
SD of ML/CL Res.	\$0.204,\$0.189	\$0.156	\$0.139	\$0.161,\$0.149	\$0.179
IQR p_{ML}, p_{CL}	\$0.222,\$0.210	\$0.273	\$0.250	\$0.228,\$0.206	\$0.316
Skewness of BL Res.	-0.361	-0.353	-0.005	-0.307	-0.314
ML/CL Res.	-0.100,-0.329	-0.331	-0.004	-0.296,-0.201	-0.297

Notes: BL = Bud Light, ML = Miller Lite and CL=Coors Light. SD = standard deviation. Res. = residuals from the AR(1) regressions. For the data we report separate values for the statistics for ML and CL, but, because the model assumes that ML and CL are symmetric, and so predicts identical statistics (ignoring simulation error), we match the average of these values during estimation and report a single prediction.

data that contains temporary price reductions. The estimated marginal cost ranges are much larger than in our examples, but the estimated σ_cs imply that the probability that a marginal cost can go from high to low across periods is quite high.[38] The CI specification in column (7) uses the same demand and data as the signaling model in column (1). As the CI model implies smaller average markups and does not tend to augment cost volatility, the marginal cost estimates and the width of the cost interval increase.

The upper panel of Table 6 reports the fit of the moments that we match during the calibration for the column (1), (2) and (7) specifications. The signaling models match all six moments accurately, but the CI model predicts that the cross-brand ρs should be approximately zero rather than positive. The lower part of the table reports moments that are not matched, including the skewness of the innovations from the AR(1) regression which is unrelated to any

[38]For example, for the specification in column (1) the probability that a firm with marginal cost $\overline{c_i}$ will have a marginal cost in the lower half of the range in the next period is 0.24, similar to 0.32 in our baseline example.

Table 7: Predicted Average Prices Before and After the MC JV For Signaling Model

	(1)	(2)	(3)	(4)	(5)	(6)
Frequency	Week	Week	Week	Week	Week	Month
Sizes	All	12 only	All	All	All	All
Price Reductions	Excl.	Excl.	Incl.	Excl.	Excl.	Excl.
Brand Elasticity	-3	-3	-3	-2.5	-3.5	-3
Flagship Diversion	85%	85%	85%	90%	80%	85%
Pre-JV Mean Prices						
BL	\$10.09	\$10.30	\$9.81	\$10.09	\$10.09	\$10.09
ML/CL	\$9.96	\$10.22	\$9.68	\$9.96	\$9.96	\$9.95
Assumed ML/CL Synergy	-\$1.18	-\$1.20	-\$1.14	-\$1.50	-\$0.94	-\$1.17
Post-JV Mean Prices						
BL	\$10.62	\$10.90	\$10.17	\$10.98	\$10.42	fails
	(+5.3%)	(+5.7%)	(+3.7%)	(+8.7%)	(+3.3%)	
ML/CL	\$10.48	\$10.79	\$10.02	\$10.82	\$10.27	fails
	(+5.2%)	(+5.8%)	(+3.5%)	(+8.5%)	(+3.1%)	

Notes: BL = Bud Light, ML = Miller Lite and CL=Coors Light. For the data we report separate values for the statistics for ML and CL, but, because the model assumes that ML and CL are symmetric, and so predicts identical statistics (ignoring simulation error), we report a single prediction.

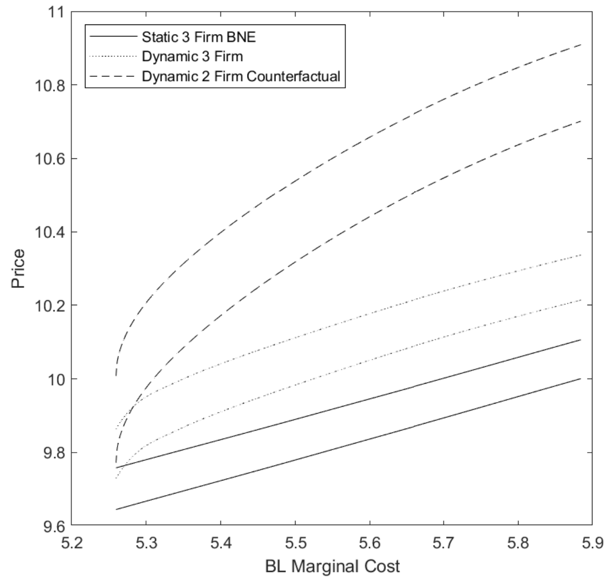
of the estimation moments. All of the models underpredict the standard deviation of price equation residuals for BL, although the difference is smaller for the specification calibrated using only 12-pack data. The signaling models match the skewness of the pricing residuals accurately, but the CI model also fails to match this dimension of the data.

4.2.6 Predicted Effects of the JV.

Table 7 reports predicted prices when we resolve the six signaling models assuming that ML and CL have the same marginal cost and are sold by a single firm at the same price. We assume that MC benefits from a synergy that would have prevented average prices from rising if firms set static CI Nash prices, as this appears consistent with the DOJ's expectation, but the width of the cost interval and the remaining parameters remain the same. The predicted price changes in columns (1)-(5) are all within the estimated ranges of 40 cents to \$1 or 3-6%.[39] We cannot find an equilibrium for the monthly data specification. In this case, the estimated parameters

[39]One might be concerned that our assumed discount factor of $\beta = 0.99$ is too low for weekly data. We have recomputed the column (1) estimates assuming $\beta = 0.998$, implying an annual discount factor of around 0.9. While a higher discount factor increases signaling incentives, the estimated parameters change to rationalize pre-JV dynamics in such a way that the predicted post-JV prices are within 1 cent of those reported in Table 7.

Figure 8: Bud Light Equilibrium Pricing Strategies (for estimates in column (1) of Table 5).



Notes: the strategies shown assume that $c_{t-1}^{BL} = \underline{c^{BL}}$ and $c_{t-1}^{ML} = c_{t-1}^{CL} = \underline{c^{ML/CL}}$ (lower line) and $c_{t-1}^{BL} = \overline{c^{BL}}$ and $c_{t-1}^{ML} = c_{t-1}^{CL} = \overline{c^{ML/CL}}$ (upper line). Therefore, for each type of equilibrium, the maximum range of BL's prices spans from the lowest point on the bottom line to the highest point on the upper line.

imply marginal costs are more persistent (the probability that a firm with the cost $\overline{c_i}$ will have a cost less than $\frac{\underline{c_i}+\underline{c_i}+c'}{2}$ is only 0.067) because, in this case, we are matching coefficients from a regression that does not control for cross-store heterogeneity in retail prices, and signaling incentives raise prices so high that the conditions for separation fail.

Figure 8 compares, using the column (1) parameters, BL's equilibrium pricing strategies for the static Bayesian Nash 3-firm model, the estimated signaling 3-firm model and the counterfactual post-JV model. Signaling increases the level and the range of BL prices, which span from the lowest point on the two BL pricing functions to the highest point, especially in the counterfactual.

Table 8 compares the cross-market averages of the IQR and ρ parameter statistics before and after the JV in the data, and the values predicted by the column (1) model. It also reports the values predicted by the CI model when we assume that, after the JV, the firms use first-order conditions with a "conduct parameter" of 0.15 on the profits of their rivals. This value leads the CI model to predict the same average post-JV prices for AB and MC as the signaling model.

Table 8: Observed and Predicted Changes in Price Dynamics for Calibrated Signaling Model (Table 5, col 1) and the Calibrated CI Model (Table 5, col 7) with a Conduct Parameter ($\theta = 0.15$) to Predict the Same Change in Average Prices.

	Data			Calibrated Signaling Model			Calibrated CI Model with Conduct		
	Pre-JV	Post-JV	Change	Pre-JV	Post-JV	Change	Pre-JV	Post-JV	Change
IQR of Prices									
BL	\$0.189	\$0.241	+0.052	\$0.189	\$0.368	+0.180	\$0.188	\$0.220	+0.032
ML	\$0.222	\$0.256	+0.034	\$0.273	\$0.369	+0.096	\$0.249	\$0.216	-0.033
CL	\$0.210	\$0.244	+0.034	\$0.273	\$0.369	+0.096	\$0.249	\$0.216	-0.033
AR(1) Regression Coefficients									
$\rho^{BL,BL}$	0.444	0.524	+0.080	0.403	0.440	+0.037	0.412	0.409	-0.002
$\rho^{ML,ML}$	0.402	0.483	+0.081	0.408	0.439	+0.031	0.407	0.413	+0.006
$\rho^{CL,CL}$	0.413	0.453	+0.040	0.408	0.439	+0.031	0.407	0.413	+0.006
$\rho^{BL,ML}$	0.082	0.092	+0.010	0.074	0.149	+0.068	-0.000	0.002	+0.003
$\rho^{BL,CL}$	0.066	0.095	+0.029	0.074	0.149	+0.068	-0.000	0.002	+0.003
$\rho^{ML,BL}$	0.059	0.087	+0.028	0.046	0.154	+0.108	0.004	-0.000	-0.005
$\rho^{CL,BL}$	0.042	0.080	+0.038	0.046	0.154	+0.108	-0.002	-0.000	+0.002

Notes: BL = Bud Light, ML = Miller Lite and CL=Coors Light. The calculation of the statistics is explained in Section 4.2.4. Pre-JV averages are calculated for 45 markets, and post-JV averages are calculated for 44 markets, as one market does not have at least 5 stores observed in consecutive weeks after the JV. The CI Model simulations use the parameter estimates from Table 5, column 7, which assumes CI Nash pricing before the JV, but that after the JV the firms use a conduct parameter of 0.15. These assumptions predict average BL and ML prices of post-JV \$10.62 and \$10.46, which are almost identical to those predicted by the signaling model for the same demand system.

The signaling model correctly predicts the sign of the changes in each of the reported statistics (i.e., there is more variation in prices, and more own-brand and cross-brand serial correlation), although it does not predict which statistic increases the most. We view the fact that our model matches the qualitative changes in dynamics, as well as the increase in average price levels, even though the parameters are calibrated using only pre-JV data, as an encouraging result. On the other hand, the CI model, which was unable to match the cross-brand ρs in the pre-JV data, predicts that the serial correlation parameters should not change and that the IQRs of prices for MC and AB should change in opposite directions.

5 Conclusion

We have developed a model where oligopolists simultaneously use prices to signal private information that is relevant for their future pricing decisions. While the theoretical literature identified more than thirty years ago that this type of behavior could raise prices, we provide the first attempt to quantify how large these effects might be, and we believe that we also provide the first empirical analysis of a simultaneous signaling model. We find that even when signals can only provide limited information about future values of firms' state variables, effects on equilibrium prices can be large, and that these effects could materially affect the analysis of horizontal mergers, as well as other applications where pricing is important. Our application shows that our model can explain observed increases in the level of domestic beer prices and changes in price dynamics after the Miller-Coors joint venture without the need to appeal to changes in the nature of equilibrium play. The model also provides a natural explanation for period-to-period price volatility observed in both the beer data and data from other industries where economists have suggested that firms act collusively (Ordover (2007)) even though standard theories of tacit collusion provide no explanation for why volatility should be observed. Our theory is also consistent with how firms treat information about their marginal costs and their margins as highly confidential. As we have shown, our analysis assumes that firms can be viewed as choosing a single price each period and that demand is sufficiently inelastic that a separating equilibrium will exist. Therefore, while we believe our analysis is useful for understanding deviations from Nash pricing in some industries, we recognize that further work is needed to investigate what may happen when these assumptions are relaxed.

We have often been asked how our model and our empirical analysis relate to theories of "coordinated effects" in merger analysis. There is no standard definition of coordinated effects: the presentation in Ordover (2007) is focused on variants of tacit collusion models, but Baker and Farrell (2020) and Farrell and Baker (2021) use a much broader definition which includes both "purposive" theories of collusion and "non-purposive" theories, such as the non-collusive CI Markov Perfect models of Maskin and Tirole (1988) which show that asynchronous price-setting can lead to price levels and price dynamics that differ from simultaneous Nash. Our work shows that small and plausible asymmetries of information can lead to similar patterns, and this may be a more plausible explanation in industries where all firms are able to change prices frequently. On a more technical level, asymmetric information can also be convenient because it means that each firm chooses its price against a distribution of its rivals' anticipated prices. This feature of asymmetric information models has long been appreciated in the static and dynamic literatures on discrete choice games (e.g., Seim (2006)), but the benefits also arise when choices are continuous.

Non-purposive theories suggest that agencies reviewing mergers, and courts, should be skeptical about relying on static CI Nash models to predict price effects in concentrated markets even when there is no evidence of collusion in an industry prior to a transaction. They can also explain why coordinated effects do not raise prices to joint-profit maximizing levels, an outcome that models of tacit collusion will usually predict when firms are even moderately patient, but which most economists believe does not happen in practice. However, future work could valuably combine signaling and collusion into a single model, building on Kreps, Milgrom, Roberts, and Wilson (1982), Athey and Bagwell (2008) and one of our examples in Appendix B.2 which illustrates how signaling could greatly magnify the effect of any small incentive that firms' have to raise their rivals' profits.

References

Ashenfelter, O., D. Hosken, and M. Weinberg (2014): "Did Robert Bork Understate the Competitive Impact of Mergers? Evidence from Consummated Mergers," *Journal of Law and Economics*, 57(S3), S67–S100.

Ashenfelter, O. C., D. S. Hosken, and M. C. Weinberg (2015): "Efficiencies Brewed: Pricing and Consolidation in the US Beer Industry," *RAND Journal of Economics*, 46(2), 328–361.

Asker, J., C. Fershtman, J. Jeon, and A. Pakes (2020): "A Computational Framework for Analyzing Dynamic Procurement Auctions: The Market Impact of Information Sharing," *RAND Journal of Economics*, 51(3), 805–839.

Athey, S., and K. Bagwell (2008): "Collusion with Persistent Cost Shocks," *Econometrica*, 76(3), 493–540.

Athey, S., K. Bagwell, and C. Sanchirico (2004): "Collusion and Price Rigidity," *Review of Economic Studies*, 71(2), 317–349.

Azar, J., M. C. Schmalz, and I. Tecu (2018): "Anticompetitive Effects of Common Ownership," *Journal of Finance*, 73(4), 1513–1565.

Baker, J. B., and J. Farrell (2020): "Oligopoly Coordination, Economic Analysis, and the Prophylactic Role of Horizontal Merger Enforcement," *University of Pennsylvania Law Review*, 168(7), 1985–2018.

Berry, S. T., and P. A. Haile (2014): "Identification in Differentiated Products Markets Using Market Level Data," *Econometrica*, 82(5), 1749–1797.

Bonatti, A., G. Cisternas, and J. Toikka (2017): "Dynamic Oligopoly with Incomplete Information," *Review of Economic Studies*, 84(2), 503–546.

Bresnahan, T. F. (1982): "The Oligopoly Solution Concept is Identified," *Economics Letters*, 10(1-2), 87–92.

——— (1989): "Empirical Studies of Industries with Market Power," *Handbook of Industrial Organization*, 2, 1011–1057.

Bronnenberg, B. J., M. W. Kruger, and C. F. Mela (2008): "The IRI Marketing Dataset," *Marketing Science*, 27(4), 745–748.

Caminal, R. (1990): "A Dynamic Duopoly Model with Asymmetric Information," *Journal of Industrial Economics*, 38(3), 315–333.

Cho, I., and D. Kreps (1987): "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics*, 102(2), 179–221.

Corts, K. S. (1999): "Conduct Parameters and the Measurement of Market Power," *Journal of Econometrics*, 88(2), 227–250.

De Angelis, D., and Y. Grinstein (2014): "Performance Terms in CEO Compensation Contracts," *Review of Finance*, 19(2), 619–651.

Doraszelski, U., and J. Jaumandreu (2013): "R&D and Productivity: Estimating Endogenous Productivity," *Review of Economic Studies*, 80(4), 1338–1383.

Ericson, R., and A. Pakes (1995): "Markov-Perfect Industry Dynamics: A Framework for Empirical Work," *Review of Economic Studies*, 62(1), 53–82.

Farrell, J., and J. B. Baker (2021): "Natural Oligopoly Responses, Repeated Games, and Coordinated Effects in Merger Analysis: A Perspective and Research Agenda," *Review of Industrial Organization*, 58(2), 1–39.

Fershtman, C., and A. Pakes (2012): "Dynamic Games with Asymmetric Information: A Framework for Empirical Work," *Quarterly Journal of Economics*, 127(4), 1611–1661.

Garmon, C. (2017): "The Accuracy of Hospital Merger Screening Methods," *RAND Journal of Economics*, 48(4), 1068–1102.

Hubbard, T. P., and H. J. Paarsch (2013): "On the Numerical Solution of Equilibria in Auction Models with Asymmetries within the Private-Values Paradigm," in *Handbook of Computational Economics*, ed. by K. Schmedders, and K. L. Judd, vol. 3, pp. 37–115. Elsevier: North-Holland.

Judd, K. L. (1998): *Numerical Methods in Economics*. MIT Press.

Katz, M. L. (1991): "Game-Playing Agents: Unobservable Contracts as Precommitments," *RAND Journal of Economics*, 22(3), 307–328.

Kaya, A. (2009): "Repeated Signaling Games," *Games and Economic Behavior*, 66(2), 841–854.

Kreps, D. M., P. Milgrom, J. Roberts, and R. Wilson (1982): "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma," *Journal of Economic Theory*, 27(2), 245–252.

Lau, L. J. (1982): "On Identifying the Degree of Competitiveness from Industry Price and Output Data," *Economics Letters*, 10(1-2), 93–99.

MacKay, A., and N. H. Miller (2019): "Estimating Models of Supply and Demand: Instruments and Covariance Restrictions," Discussion paper, Harvard University.

Mailath, G. J. (1987): "Incentive Compatibility in Signaling Games with a Continuum of Types," *Econometrica*, 55(6), 1349–1365.

Mailath, G. J. (1988): "An Abstract Two-Period Game with Simultaneous Signaling: Existence of Separating Equilibria," *Journal of Economic Theory*, 46(2), 373–394.

——— (1989): "Simultaneous Signaling in an Oligopoly Model," *Quarterly Journal of Economics*, 104(2), 417–427.

Mailath, G. J., and E.-L. von Thadden (2013): "Incentive Compatibility and Differentiability: New Results and Classic Applications," *Journal of Economic Theory*, 148(5), 1841–1861.

Maskin, E., and J. Tirole (1988): "A Theory of Dynamic Oligopoly, II: Price Competition, Kinked Demand Curves, and Edgeworth Cycles," *Econometrica*, 56(3), 571–599.

Mester, L. J. (1992): "Perpetual Signalling with Imperfectly Correlated Costs," *RAND Journal of Economics*, 23(4), 548–563.

Milgrom, P., and J. Roberts (1982): "Limit Pricing and Entry Under Incomplete Information: An Equilibrium Analysis," *Econometrica*, 50(2), 443–459.

Miller, N. H., G. Sheu, and M. C. Weinberg (2020): "Oligopolistic Price Leadership and Mergers: The United States Beer Industry," Discussion paper, Georgetown University.

Miller, N. H., and M. C. Weinberg (2017): "Understanding the Price Effects of the MillerCoors Joint Venture," *Econometrica*, 85(6), 1763–1791.

Murphy, K. J. (1999): "Executive Compensation," *Handbook of Labor Economics*, 3, 2485–2563.

Nevo, A. (1998): "Identification of the Oligopoly Solution Concept in a Differentiated-Products Industry," *Economics Letters*, 59(3), 391–395.

Nocke, V., and M. D. Whinston (2020): "Concentration Screens for Horizontal Mergers," Discussion Paper 27533, National Bureau of Economic Research.

O'Brien, D. P., and S. C. Salop (1999): "Competitive Effects of Partial Ownership: Financial Interest and Corporate Control," *Antitrust Law Journal*, 67, 559–614.

Olley, G. S., and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64(6), 1263–1297.

Ordover, J. A. (2007): "Coordinated Effects in Merger Analysis: An Introduction," *Columbia Business Law Review*, (2), 411–436.

Pakes, A., and P. McGuire (1994): "Computing Markov-Perfect Nash Equilibria: Numerical Implications of a Dynamic Differentiated Product Model," *RAND Journal of Economics*, 25(4), 555–589.

Peters, C. (2009): "Evaluating the Performance of Merger Simulation: Evidence from the U.S. Airline Industry," *Journal of Law and Economics*, 49(2), 627–649.

Porter, R. H. (1983): "A Study of Cartel Stability: The Joint Executive Committee, 1880-1886," *Bell Journal of Economics*, 14(2), 301–314.

Roddie, C. (2012): "Signaling and Reputation in Repeated Games, I: Finite Games," Discussion paper, University of Cambridge.

Rotemberg, J. J., and D. S. Scharfstein (1990): "Shareholder-Value Maximization and Product-Market Competition," *Review of Financial Studies*, 3(3), 367–391.

Seim, K. (2006): "An Empirical Model of Firm Entry with Endogenous Product-Type Choices," *RAND Journal of Economics*, 37(3), 619–640.

Shapiro, C. (1986): "Exchange of Cost Information in Oligopoly," *Review of Economic Studies*, 53(3), 433–446.

Sklivas, S. D. (1987): "The Strategic Choice of Managerial Incentives," *RAND Journal of Economics*, 18(3), 452–458.

Smith, A. (2008): "Indirect Inference," *The New Palgrave Dictionary of Economics, 2nd Edition.*

Sweeting, A., J. W. Roberts, and C. Gedge (2020): "A Model of Dynamic Limit Pricing with an Application to the Airline Industry," *Journal of Political Economy*, 128(3), 1148–1193.

Toxvaerd, F. (2008): "Strategic Merger Waves: A Theory of Musical Chairs," *Journal of Economic Theory*, 140(1), 1–26.

——— (2017): "Dynamic Limit Pricing," *RAND Journal of Economics*, 48(1), 281–306.

Vives, X. (2011): "Strategic Supply Function Competition with Private Information," *Econometrica*, 79(6), 1919–1966.

APPENDICES TO "DYNAMIC OLIGOPOLY PRICING WITH ASYMMETRIC INFORMATION: IMPLICATIONS FOR HORIZONTAL MERGERS" FOR ONLINE PUBLICATION

A Computational Algorithms

This Appendix describes the methods used to solve our model. We describe the continuous type, finite horizon model in detail, before noting what changes in other cases. Our discussion will assume that there are two ex-ante symmetric duopolists. When firms are asymmetric, all of the operations need to be repeated for each firm.

A.1 Finite Horizon Model.

A.1.1 Preliminaries.

We specify discrete grids for the actual and perceived marginal costs of each firm, which will be used to keep track of expected per-period profits, value functions and pricing strategies. For example, when each firm's marginal cost lies on $[8, 8.05]$ and we use 8-point equally spaced grids, the points are $\{8,\ 8.0071,\ 8.0143,\ 8.0214,\ 8.0286,\ 8.0357,\ 8.0429,\ 8.0500\}$.[40] We use interpolation and numerical integration to account for the fact that realized types will lie between these isolated points. The discount factor is $\beta = 0.99$.

It is useful to define several functions that we will use below:

- $P_{i,t}\left(\widehat{c^{j}_{i,t-1}}, c_{j,t-1}\right)$ is firm i's pricing function in period t. This is a function of the marginal cost that j believes that i had in the previous period, $\widehat{c^{j}_{i,t-1}}$ (which, when j is forming equilibrium beliefs, will reflect that cost that i signaled in the previous period). It will also depend on the marginal cost that i believes that j had in the previous period, but we solve the game assuming that j is using its equilibrium strategy, so that i assumes that its perception of j's prior cost is correct, so we use the argument $c_{j,t-1}$. The actual price set will depend on $c_{i,t}$, and, when we need to integrate over the values that $p_{i,t}$ may

[40] The examples reported in Section 3 use 12 gridpoints, although we have experimented with as many as 20 gridpoints in each dimension to make sure that this does not have a material effect on the reported results.

take (e.g., to calculate expected profits) we will include $c_{i,t}$ as an explicit argument in the function.

- $\pi_i(p_{i,t}, p_{j,t}, c_{i,t})$ is firm i's one-period profit when it has marginal cost $c_{i,t}$ and sets price $p_{i,t}$, and its rival sets price $p_{j,t}$. This function does not depend on t because demand is assumed to be static and time-invariant.

- $V_{i,t}\left(c_{i,t-1}, \widehat{c_{i,t-1}^{j}}, c_{j,t-1}\right)$ is the value function for firm i defined at the beginning of period t, before firm types have evolved to their period t values. It reflects the expected payoffs of firm i in period t and the discounted value of expected payoffs in future periods given equilibrium play in both t and future periods. It depends on the true value of each firm's type in $t-1$, and the rival's perception of i's $t-1$ type (reflecting any deviation that i made in $t-1$). In the case of an 8-point grid, $V_{i,t}$ is a 512x1 vector.

- $\Pi^{i,t}\left(c_{i,t}, \widehat{c_{i,t}^{j}}, p_{i,t}, \widehat{c_{i,t-1}^{j}}, c_{j,t-1}\right)$ is the intermediate signaling payoff function of firm i when it knows its current marginal cost $c_{i,t}$, and is deciding what price to set. It does not know the period t type of its rival, but it reflects the pricing function that i expects j to use, $P_{j,t}\left(c_{j,t-1}, \widehat{c_{i,t-1}^{j}}\right)$. $\widehat{c_{i,t}^{j}}$ is the perception that j will have about i's cost at the end of period t. When the rival sets price $P_{j,t}\left(c_{j,t}, c_{j,t-1}, \widehat{c_{i,t-1}^{j}}\right)$,

$$\Pi^{i,t}\left(c_{i,t}, \widehat{c_{i,t}^{j}}, p_{i,t}, \widehat{c_{i,t-1}^{j}}, c_{jt-1}\right) = \int_{\underline{c}_j}^{\overline{c}_j} \left(\begin{array}{c} \pi_i\left(p_{i,t}, P_{j,t}\left(c_{j,t}, c_{j,t-1}, \widehat{c_{i,t-1}^{j}}\right), c_{i,t}\right) + \\ \beta V_{i,t+1}\left(c_{i,t}, \widehat{c_{i,t}^{j}}, c_{j,t}\right) \end{array} \right) \psi_j(c_{j,t}|c_{j,t-1}) dc_{j,t}.$$

where we note that $p_{i,t}$ only enters through current profits, and $\widehat{c_{i,t}^{j}}$ only enters through the discounted continuation value. In practice, our description will make up $\Pi^{i,t}$ into two components: $\Pi^{i,t} = \widetilde{\pi}_i + \widetilde{V_{i,t}}$, where

$$\widetilde{\pi}_i\left(p_{i,t}, P_{j,t}\left(c_{j,t-1}, \widehat{c_{i,t-1}^{j}}\right), c_{i,t}\right) = \int_{\underline{c}_j}^{\overline{c}_j} \pi_i\left(p_{i,t}, P_{j,t}\left(c_{j,t}, c_{j,t-1}, \widehat{c_{i,t-1}^{j}}\right), c_{i,t}\right) \psi_j(c_{j,t}|c_{j,t-1}) dc_{j,t}$$

and

$$\widetilde{V_{i,t}}\left(c_{i,t},\widehat{c_{i,t}^{j}},c_{j,t-1}\right)=\int_{\underline{c}_j}^{\bar{c}_j}\beta V_{i,t+1}\left(c_{i,t},\widehat{c_{i,t}^{j}},c_{j,t}\right)\psi_j(c_{j,t}|c_{j,t-1})dc_{j,t}.$$

Given a set of fully separating pricing functions $P_{i,t}\left(\widehat{c_{i,t-1}^{j}},c_{j,t-1}\right)$, the relationship between Π and V is that

$$V_{i,t}\left(c_{i,t-1},\widehat{c_{i,t-1}^{j}},c_{j,t-1}\right)=\int_{\underline{c}_i}^{\bar{c}_i}\Pi^{i,t}\left(c_{i,t},c_{i,t},P_{i,t}\left(c_{i,t},\widehat{c_{i,t-1}^{j}},c_{j,t-1}\right),\widehat{c_{i,t-1}^{j}},c_{j,t-1}\right)\psi_i(c_{i,t}|c_{i,t-1})dc_{i,t}$$

where we recognize that, in equilibrium, i's period t pricing function will reveal its cost to j, implying $\widehat{c_{i,t}^{j}}=c_{i,t}$.

A.1.2 Period T.

Assuming that play in period $T-1$ was fully separating, we solve for BNE pricing strategies for each possible combination of beliefs (on our grid) about period $T-1$ marginal costs. A strategy for each firm is an optimal price given the realized value of its own period T cost, given the pricing strategy of the rival, its prior marginal cost and the rival's belief about the firm's period $T-1$ cost. Trapezoidal integration is used to integrate over the realized cost/price of the rival using a discretized version of the pdf of each firm's cost transition, and we solve for the BNE prices using the implied first-order conditions (i.e., those associated with maximizing static profits). With symmetric duopolists and 8-point grids, we find 512 equilibrium prices.

We use the equilibrium prices to calculate the beginning of period value function

$$V_{i,T}\left(c_{i,T-1},\widehat{c_{i,T-1}^{j}},c_{j,T-1}\right)=...$$

$$\int_{\underline{c}_i}^{\bar{c}_i}\int_{\underline{c}_j}^{\bar{c}_j}\pi_i\left(P_{i,T}^{*}\left(c_{i,T},\widehat{c_{i,T-1}^{j}},c_{j,T-1}\right),P_{j,T}^{*}\left(c_{j,T},c_{j,T-1},\widehat{c_{i,T-1}^{j}}\right),c_{i,T}\right)\psi_j(c_{j,T}|c_{j,T-1})\psi_i(c_{i,T}|c_{i,T-1})dc_{j,T}dc_{i,T}.$$

A.1.3 Period $T-1$.

Firms choose prices in period $T-1$ recognizing that their prices will affect rivals' prices in period T. We solve for period $T-1$ strategies, assuming separating equilibrium pricing and interpretation of beliefs in period $T-2$, so that each firm has a point belief about its rival's period $T-2$ marginal cost. We then use the following steps to compute equilibrium strategies.

Step 1. (a) Compute

$$\widetilde{V}_{i,T-1}\left(c_{i,T-1}, \widehat{c^j_{i,T-1}}, c_{j,T-2}\right) = \beta \int_{\underline{c}_j}^{\overline{c}_j} V_{i,T}\left(c_{i,T-1}, \widehat{c^j_{i,T-1}}, c_{j,T-1}\right)\psi_j(c_{j,T-1}|c_{j,T-2})dc_{j,T-1}.$$

$\widetilde{V}_{i,T-1}$ is the expected continuation value (i.e., not including period $T-1$ payoffs) for i when it is setting its period $T-1$ price, without knowing the period $T-1$ realization of c_j (but knowing that, in equilibrium, it will be revealed by $p_{j,T-1}$).

(b) Compute $\beta \frac{\partial \widetilde{V}_{i,T-1}\left(c_{i,T-1}, \widehat{c^j_{i,T-1}}, c_{j,T-2}\right)}{\partial \widehat{c^j_{i,T-1}}}$ using numerical differences at each of the gridpoints (one-sided as appropriate). This array provides us with a set of values for the numerator in the differential equation (1). These derivatives do not depend on period $T-1$ prices, so we do not repeat this calculation as we look for equilibrium strategies.

(c) Verify belief monotonicity using these derivatives.

Step 2. We use the following iterative procedure to solve for equilibrium fully separating prices.[41] Use the BNE prices (i.e., those calculated in period T) as initial starting values. Set the iteration counter, $iter = 0$.

(a) Given the current guess of the strategy of firm j, $P_{j,T-1}\left(c_{j,T-1}, c_{j,T-2}, \widehat{c^j_{i,T-2}}\right)$, which is equal to the pricing functions solved for in the previous iteration, calculate $\frac{\partial \widetilde{\pi_{i,T-1}}\left(p_{i,T-1}, P_{j,T-1}\left(c_{j,T-2}, \widehat{c^j_{i,T-2}}\right), c_{i,T-1}\right)}{\partial p_{i,T-1}}$ for a grid of values $\left(p_{i,T-1}, \widehat{c^j_{i,T-2}}, c_{i,T-1}\right)$ where

$$\widetilde{\pi_{i,T-1}}\left(p_{i,T-1}, P_{j,T-1}\left(c_{j,T-2}, \widehat{c^j_{i,T-2}}\right), c_{i,T-1}\right) =$$

$$\int_{\underline{c}_j}^{\overline{c}_j} \pi_i\left(p_{i,T-1}, P_{j,T-1}\left(c_{j,T-1}, c_{j,T-2}, \widehat{c^j_{i,T-2}}\right), c_{i,T-1}\right)\psi_j(c_{j,T-1}|c_{j,T-2})dc_{j,T-1}$$

[41] We do not claim that this iterative procedure is computationally optimal, although it works reliably in our examples. There are some parallels between our problem and the problem of solving for equilibrium bid functions in asymmetric first-price auctions where both the lower and upper bounds of bid functions are endogenous. Hubbard and Paarsch (2013) provide a discussion of the types of methods that are used for these problems.

i.e., the derivative of i's expected profit with respect to its price, given that it does not know what price j will charge because it does not know $c_{j,T-1}$. The derivatives are evaluated on a fine grid (steps of one cent) of prices.[42] This vector will be used to calculate the denominator in the differential equation (1).

For each $\left(\widehat{c_{i,T-2}^{j}}, c_{j,T-2}\right)$,

(b) Solve the lower boundary condition equation $\frac{\partial\widetilde{\pi}\left(p_{i,T-1}^{*}, P_{j,T-1}\left(c_{j,T-1}, c_{j,T-2}, \widehat{c_{i,T-2}^{j}}\right), \underline{c_i}\right)}{\partial p_{i,T-1}} = 0$ for $p_{i,T-1}^{*}$, using a cubic spline to interpolate the vector calculated in (a). This gives the static best response price and the lowest price on i's pricing function.

(c) Using this price as the initial point[43], solve the differential equation, (1), to find i's best response signaling pricing function. This is done using `ode113` in MATLAB, with cubic spline interpolation used to calculate the values of the numerator and the denominator between the gridpoints.[44] Interpolation is then used to calculate values for the pricing function for the specific values of $c_{i,T-1}$ on the cost/belief grid $\left(c_{i,T-1}, \widehat{c_{i,T-2}^{j}}, c_{j,T-2}\right)$.

(d) Update the current guess of i's pricing strategy using

$$P_{i,T-1}^{iter=k+1}\left(c_{i,T-1}, \widehat{c_{i,T-2}^{j}}, c_{j,T-2}\right) = (1-\tau)P_{i,T-1}^{iter=k}\left(c_{i,T-1}, \widehat{c_{i,T-2}^{j}}, c_{j,T-2}\right) + ...$$
$$\tau P_{i,T-1}^{'}\left(c_{i,T-1}, \widehat{c_{i,T-2}^{j}}, c_{j,T-2}\right) \quad \forall c_{i,T-1}, \widehat{c_{i,T-2}^{j}}, c_{j,T-2}$$

where $P_{i,T-1}^{'}$ are the best response functions that have just been computed. In the finite horizon case, $\tau = 1$, i.e., full updating, works effectively unless we are close to prices where the conditions required to characterize the unique best response fail to hold, in which case we also try using $\tau = \frac{1}{1+iter^{\frac{1}{6}}}$. See discussion below for how we update in the application where we use an infinite horizon model.

[42] A fine grid is required because it is important to evaluate the derivatives accurately around the static best response, where the derivative will be equal to zero.

[43] In practice, the exact value of the derivative will be zero at the static best response, so that the differential equation will not be well-defined if this derivative is plugged in. We therefore begin solving the differential equation at the price where $\Pi_3^{i,T-1} + 1e - 4 = 0$. Pricing functions are essentially identical if we add 1e-5 or 1e-6 instead.

[44] See discussion of tolerances in Appendix A.2.2.

(e) Check if the maximum difference between $P_{i,T-1}^{iter=k}$ and $P_{i,T-1}^{'}$, across all gridpoints, is less than 1e-6. If so, terminate the iterative process, else update the iteration counter to $iter = iter + 1$, and return to step 2(a).

(f) Verify that the solved pricing functions are monotonic in a firm's own marginal costs, and that, given the pricing functions of the rival, that the single-crossing condition holds for the full range of prices used in the putative equilibrium.

Step 4. Compute i's value $V_{i,T-1}$,

$$V_{i,T-1}\left(c_{i,T-2}, \widehat{c_{i,T-2}^{j}}, c_{j,T-2}\right) = ...$$
$$\int_{\underline{c}_i}^{\overline{c}_i}\int_{\underline{c}_j}^{\overline{c}_j}\left\{\begin{array}{c} \pi\left(P_{i,T-1}^{*}\left(c_{i,T-1}, \widehat{c_{i,T-2}^{j}}, c_{j,T-2}\right), P_{j,T-1}^{*}\left(c_{j,T-1}, c_{j,T-2}, \widehat{c_{i,T-2}^{j}}\right), c_{i,T-1}\right) \\ +\beta V_{i,T}\left(c_{i,T-1}, c_{j,T-1}, c_{i,T-1}\right) \end{array}\right\} * ...$$
$$\psi_j(c_{j,T-1}|c_{j,T-2})\psi_i(c_{i,T-1}|c_{i,T-2})dc_{j,T-1}dc_{i,T-1}$$

where we are recognizing that equilibrium play at period $T-1$ will reveal i's true cost to j. Note that this is the case even if, hypothetically, $\widehat{c_{i,T-2}^{j}} \neq c_{i,T-2}$ (i.e., j was misled in period $T-2$) because i should find it optimal to use its equilibrium signaling strategy given its new cost $c_{i,T-1}$ in response to j using a strategy based on its $\widehat{c_{i,T-2}^{j}}$ belief.

A.1.4 Earlier Periods.

This process is then repeated for earlier periods, with an appropriate changing of subscripts. Given our assumption that first period beliefs reflect actual costs in a fictitious prior period, this procedure will also calculate strategies in the first period of the game.

A.2 Infinite Horizon Model.

We use an infinite horizon model for some of our examples and the empirical application. We find equilibrium pricing functions in the continuous type model using a modification of the procedure described above: in particular, we follow the logic of policy function iteration (Judd (1998)) to calculate values given a set of strategies.

The equilibrium objects that we need to solve for are a set of stationary pricing functions, $P_i^*\left(\widehat{c_{i,t-1}^j}, c_{j,t-1}\right)$ and value functions $V_i\left(c_{i,t-1}, \widehat{c_{i,t-1}^j}, c_{j,t-1}\right)$ which are consistent with each other given the static profit function and the transition functions for firm types.

We start by solving the period $T-1$ game described previously (i.e., assuming that there is a one more period of play where firms will use static Bayesian Nash Equilibrium strategies) to give an initial set of signaling pricing functions ($P_i^{*,iter=1}$). We then calculate firm values in each state $\left(c_{i,t-1}, \widehat{c_{i,t-1}^j}, c_{j,t-1}\right)$ if these pricing functions were used in every period of an infinite horizon game. This is done by creating a discretized form of the state transition process and calculating

$$\widehat{V}_i^{iter=1} = [I - \beta T]^{-1} \pi_i' \left(c_{i,t-1}, \widehat{c_{i,t-1}^j}, c_{j,t-1}\right)$$

where

$$\pi_i'\left(c_{i,t-1}, \widehat{c_{i,t-1}^j}, c_{j,t-1}\right) = \int_{\underline{c}_i}^{\overline{c}_i}\int_{\underline{c}_j}^{\overline{c}_j} \left\{ \pi_i \left(\begin{array}{c} P_i^{*,iter=1}\left(c_{i,t}, \widehat{c_{i,t-1}^j}, c_{j,t-1}\right), \\ P_j^{*,iter=1}\left(c_{j,t}, c_{j,t-1}, \widehat{c_{i,t-1}^j}\right), c_{i,t} \end{array} \right) \right\} \psi_j(c_{j,t}|c_{j,t-1}) \psi_i(c_{i,t}|c_{i,t-1}) dc_{j,t} dc_{i,t}$$

and T is a transition matrix that reflects the transition probabilities for both firms' types and the behavioral assumption that equilibrium play in t (and future periods) will reveal period t costs. $P_j^{*,iter=1}\left(c_{j,t-1}, \widehat{c_{i,t-1}^j}\right)$ will reflect $P_i^{*,iter=1}$, applied to the states of the rival, when the firms are symmetric.

$\widehat{V}_i^{iter=1}$ is then used to compute a new set of pricing functions, $P_i^{*,iter=2}$, and the process is repeated until prices converge (tolerance 1e-4). Even though policy function iteration procedures do not necessarily converge, we find they work very well in our setting, when the conditions for separation hold, although it is sometimes necessary to update the pricing function to be a linear combination of the previous guess and the newly calculated best response. As illustrated in Figure 3, converged pricing functions found by this method are essentially identical to the pricing functions found for the early periods of long finite horizon games where the exact value of t has almost no effect on equilibrium pricing strategies. The computational advantage of this procedure comes from the fact that we do not perform the iterative procedure described above for every period of the game: instead there is a single iterative procedure where we solve for a single set of pricing strategies for the entire game.

A.2.1 Speeding Up Solutions By Interpolating Pricing Functions.

When we consider more than two firms and allow for asymmetries, the solution algorithm laid out above becomes slow, with most of the time spent solving differential equations. For example, with 8-point cost/belief grids, three asymmetric firms and 50 iterations, we would have to solve 25,600 differential equations. This would make estimation of the model using a nested fixed point procedure very slow. On the other hand, reducing the number of gridpoints can lead to inaccurate calculations of expected payoffs, and therefore strategies.

Examination of the equilibrium pricing functions (see, for example, Figure 3) shows that as we vary rivals' prior types, a firm's pricing functions look like they are translated without (noticeably) changing shape. We exploit this fact by solving for pricing functions for only a subset of the $\left(\widehat{c_{i,t-1}^{j}}, c_{j,t-1}\right)$ gridpoints and using cubic splines to interpolate the remaining values.[45] This allows us to achieve a substantial speed increase, while continuing to calculate expected values accurately on a finer grid.

A.2.2 Tolerances and Updating Rules Used for the Estimation of the Cost Parameters Using the Infinite Horizon Model.

In Section 4 we estimate the cost parameters using a nested fixed point algorithm, which means that both speed and accuracy are important. After considerable experimentation, we use the following tolerances:

- for the parameter search using `fminsearch` we set the tolerance for the parameter values at 1e-5 and the tolerance on changes to the objective function at 1e-5. The value of the minimized objective function is typically less than 0.0002, compared with the initial guess, for which we use estimates of the parameters assuming firms use static Bayesian Nash pricing strategies, which usually gives an objective function value of around 0.2.

- the tolerance for criterion for the pricing functions when solving the model is 1e-6 (i.e., at none of the grid points should the price on the best response pricing function be more than 1e-6 from the current guess).

[45]For example, when we estimate our model in Section 4, we use a seven-point cost grid ({1,..,7}) for the profits and values of each firm. We solve for pricing functions for the full interaction of gridpoints {1,3,5,7} and then interpolate the pricing functions for the remaining gridpoints.

- for the differential equation solver, the initial step size is 5e-5 and the maximum step size is 0.003 for the first ten iterations of the algorithm, but we then use an initial step size of 1e-5 and a maximum step size of 0.001.

- we update the pricing function to be the best response for the first 15 iterations, and then use a linear combination of the best response and the current guess where the weight on the best response changes linearly from 1 (iteration 16) to 0.1 (iteration 115).

When we use these tolerances, the infinite horizon game is typically solved using somewhere between 12 and 45 iterations, taking between 3 and 20 minutes. Estimation of the five parameters usually requires around 250 function evaluations, although the objective function and parameters are usually close to their final values within 100 evaluations.

A.3 Two-Type Model.

We use a model where each firm can have one of two types when we want to examine all strategies simultaneously or to consider a large number of alternative demand parameters. An additional advantage is that because prices, profits and values can be calculated for each possible type, we avoid small inaccuracies that result from numerical integration.

The key difference to the solution algorithm is that we no longer solve differential equations to find best response pricing functions. Recall that in the continuous type model, the differential equations characterize the unique separating best response when the signaling payoff function satisfies several conditions. In the discrete type model, one can construct multiple separating pricing functions that can be supported for different beliefs of the rival firm. To proceed we therefore need to choose a particular pricing function. We describe our choice, and the method we use to calculate the best response prices here. This procedure can be embedded within the procedure for solving either a finite horizon or an infinite horizon game.

To be as consistent with the continuous type model as possible, we use the prices that allow the two types to separate at the lowest cost, in terms of foregone current profits taking the current guess of the pricing function of the rival as given, to the signaling firm (i.e., "Riley" signaling strategies, which would also be those that satisfy application of the intuitive criterion).[46]

[46] Of course, in the game we are considering it could be advantageous to the firms to use higher signaling

The amended computational procedure is as follows (described for the infinite horizon case). Suppose that we are looking to find the pricing strategy of firm i in period t when it believes that j's previous cost was $c_{j,t-1}$ and j believes that i's previous cost was $\widehat{c^j_{i,t-1}}$. We will repeat this process for each $\left(\widehat{c^j_{i,t-1}}, c_{j,t-1}\right)$ combination, of which there will be four in the duopoly model. We need to solve for two prices: i's price when its cost is $\underline{c_i}$ and its price when its cost is $\overline{c_i}$.

Step 1. Find $p^*_{i,t}(\underline{c_i})$, which will be the static best response, as the solution to $\frac{\partial \widetilde{\pi}\left(p_{i,t}, P_{j,t}\left(c_{j,t}, c_{j,t-1}, \widehat{c^j_{i,t-1}}\right), \underline{c_i}\right)}{\partial p_{i,t}} = 0$ where

$$\widetilde{\pi}_i\left(p_{i,t}, P_{j,t}\left(c_{j,t}, c_{j,t-1}, \widehat{c^j_{i,t-1}}\right), \underline{c_i}\right) =$$
$$\pi_i\left(p_{i,t}, P_{j,t}\left(\underline{c_j}, c_{j,t-1}, \widehat{c^j_{i,t-1}}\right), \underline{c_i}\right) \Pr(c_{j,t} = \underline{c_j}|c_{j,t-1}) + ...$$
$$\pi_i\left(p_{i,t}, P_{j,t}\left(\overline{c_j}, c_{j,t-1}, \widehat{c^j_{i,t-1}}\right), \underline{c_i}\right) \Pr(c_{j,t} = \overline{c_j}|c_{j,t-1})$$

Step 2. Find $p^*_{i,t}(\overline{c_i})$. This is done by finding the price, p', higher than $p^*_{i,t}(\underline{c_i})$, which would make the low cost firm indifferent between setting price $p^*_{i,t}(\underline{c_i})$ and being perceived as a low cost type, and setting price p' and being perceived as a high cost type, i.e.,

$$\widetilde{\pi}\left(p^*_{i,t}(\underline{c_i}), P_{j,t}\left(c_{j,t}, c_{j,t-1}, \widehat{c^j_{i,t-1}}\right), \underline{c_i}\right) + \beta\widetilde{V_{i,t+1}}(\underline{c_i}, \underline{c_i}, c_{j,t-1}) = ...$$
$$\widetilde{\pi}\left(p', P_{j,t}\left(c_{j,t}, c_{j,t-1}, \widehat{c^j_{i,t-1}}\right), \underline{c_i}\right) + \beta\widetilde{V_{i,t+1}}(\underline{c_i}, \overline{c_i}, c_{j,t-1})$$

where

$$\beta\widetilde{V_{i,t+1}}\left(c_{i,t-1}, \widehat{c^j_{i,t-1}}, c_{j,t-1}\right) = V_{i,t+1}\left(c_{i,t-1}, \widehat{c^j_{i,t-1}}, \underline{c_j}\right) \Pr(c_{j,t} = \underline{c_j}|c_{j,t-1}) + ...$$
$$V_{i,t+1}\left(c_{i,t-1}, \widehat{c^j_{i,t-1}}, \overline{c_j}\right) (1 - \Pr(c_{j,t} = \underline{c_j}|c_{j,t-1})).$$

We verify that, consistent with single-crossing, the $\overline{c_i}$ type prefers to set the price p' rather than

prices, because of how this raises rivals' prices in equilibrium. This equilibrium consideration is ignored when selecting the Riley best response.

setting its static best response price. We also verify belief monotonicity when we calculate the value functions. As illustrated in Section B.1, there are parameters for which belief monotonicity fails.

B Additional Examples.

B.1 Two-Type Examples.

A model where each firm can have one of two types has a much lower computational burden than the continuous type model. In this Appendix we will consider several parameterizations of a two-type model. In all of them we assume that firms are symmetric and that in any period $c_i = \underline{c} = 8$ or $c_i = \overline{c} = 8.05$. The probability that the cost remains the same as in the last period is $0.5 \leq \rho < 1$. There are no signaling incentives when $\rho = 0.5$.

Refinement. A disadvantage of the two-type model is that for a given pricing strategy of firm j, firm i's separating best response pricing function is not unique in the sense that it depends on how firm j will interpret the signal. We therefore impose a refinement that is consistent with the logic of the "intuitive criterion" (Cho and Kreps (1987)), which has often been applied as a refinement in discrete-type signaling games where only one player is signaling. Specifically, we assume that the low cost type's strategy will be the static best response, as in the continuous type model, and, under assumptions that appropriately map Conditions 1-4 to the two-type case, the high cost type's best response price will be the lowest price that the low cost type would be unwilling to set even if this would result in rivals' perceiving it as a high cost type rather than a low cost type. While this does uniquely define the best response, it does not guarantee a unique equilibrium in the oligopoly signaling game, and we have identified several examples in the infinite horizon version of the two-type model where there are multiple equilibria. The results reported in this Appendix use an algorithm which, when an infinite horizon equilibirum exists, appears consistently to select the equilibrium which corresponds to the equilibrium in the early periods of a long finite horizon game.

Method. See Appendix A.3 for a description of the method used to solve the two-type model.

B.1.1 Outcomes for Alternative Serial Correlation and Demand Parameters.

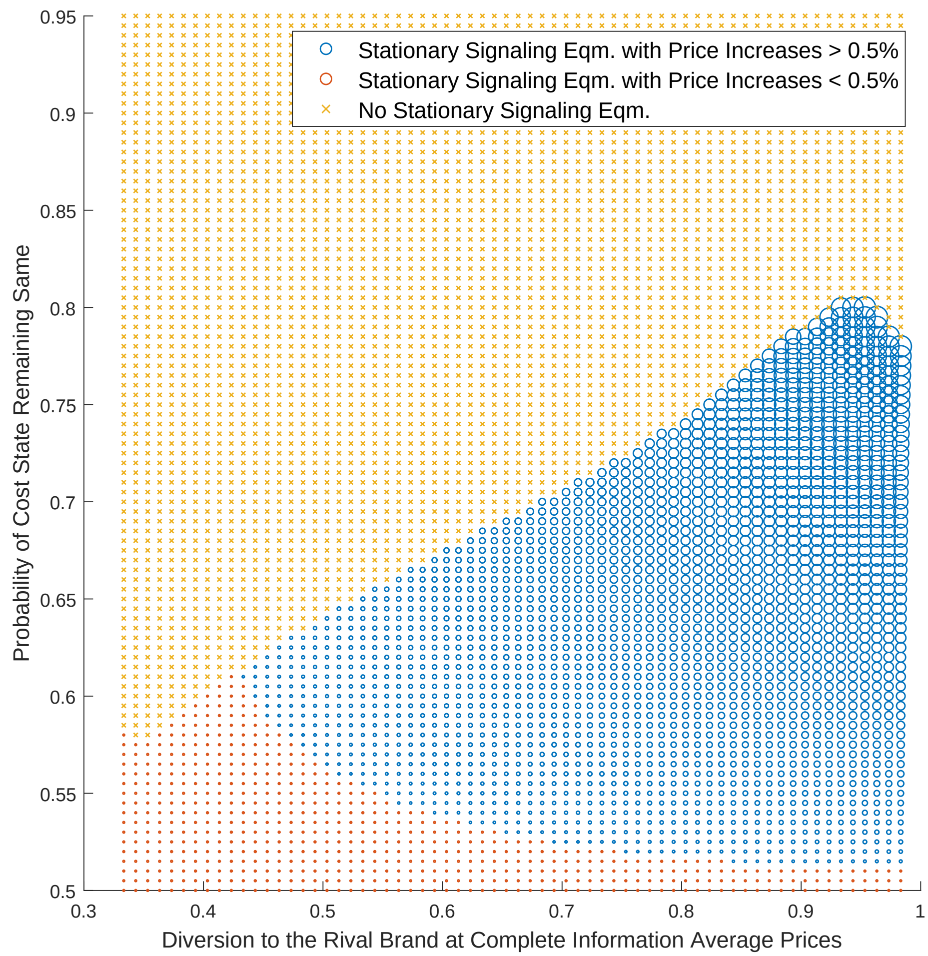
We assume nested logit demand where the indirect utility function for consumer c has the form $u_{i,c} = \beta - \alpha p_i + \sigma \nu_c + (1 - \sigma)\epsilon_{i,c}$. We choose β, α and σ so that, for each combination of parameters that we consider, the CI equilibrium prices (at average cost levels) are \$16 for each firm, the market share of each firm at these prices is 0.25, and the diversion, which measures the proportion of a product's lost demand that goes to the rival's product, rather than the outside good, when its price increases from the CI equilibrium price, has a value that we specify. We focus on diversion because when more demand goes to the outside good, which is like a competitor that always offers a fixed utility and does not respond to a signal, firms have less incentive to signal and, as we will show, the belief monotonicity and single-crossing conditions become harder to satisfy.[47] Given assumed market shares, the lowest possible value of this diversion measure is $\frac{1}{3}$, which corresponds to multinomial logit demand. We vary ρ from 0.5 (in which case there is no incentive to signal) to 0.99. We solve an infinite horizon version of our model.

Figure B.1 shows the results for a fine grid of values of diversion and ρ. The orange crosses indicate combinations where the conditions for characterizing best responses fail and we cannot find a separating equilibrium. For combinations where we can find a separating equilibrium the size and color of the circles indicate the percentage increase in average prices relative to average static Bayesian Nash equilibrium prices with the same demand and serial correlation parameters (these prices are also always very close to \$16). When serial correlation is very low, the price effects are always small whatever the level of diversion, and, for given diversion, the price effects become larger as serial correlation increases. For given serial correlation, higher diversion is associated with larger price effects, as it becomes more beneficial for a firm to increase its rival's price (because more of the demand that the rival loses will come to the firm), and the increase in a rival's price has a greater effect on the firm's best response. For moderate diversion, such as 0.6, an equilibrium cannot be sustained once serial correlation increases above 0.66. When diversion to rival products is very high, equilibria can be sustained with very large price effects: we find a maximum price increase of 44.8%.[48]

[47]The intuition is that when the rival's expected price increases, a firm may have a greater incentive to lower its price, towards a static best response price, to take demand from the outside good. See below for an example.

[48]In the diagram, the highest serial correlation for which we can find an equilibrium falls when we increase diversion above 0.95. This appears to reflect the fact that, at this level, small increases in diversion can increase signaling prices significantly, leading the conditions to fail. For each considered value of diversion above 0.95, we identify a value of ρ where signaling raises prices by more than 43.0% and 44.8%.

Figure B.1: Equilibrium Average Price Increases in the Infinite Horizon Two-Type Duopoly Model as a Function of Diversion and Serial Correlation of Costs



Notes: red dots mark outcomes where there is a stationary separating equilibrium with average prices less than 0.5% above static BNE levels. The blue circles mark outcomes where there is a stationary separating equilibrium with larger average price increases relative to static BNE prices, and the size of the circle is linearly increasing in the percentage difference in prices (the largest effect shown has average prices increasing by 44.8%). Orange crosses mark outcomes where the conditions required to solve for best response functions fail and we cannot find an equilibrium. The diversion is measured by the proportion of demand that goes to the rival product when one product experiences a small increase in price at CI Nash equilibrium prices given average costs.

B.1.2 Failure of the Conditions Required for Existence of a Separating Equilibrium.

We now consider in more detail an example where the conditions required for separation fail. Demand is the same as before (i.e., indirect utility is $u_{i,c} = 5-0.1p_i+0.25\nu_c+(1-0.25)\epsilon_{i,c}$), and each firm's marginal cost is either 8 (low) or 8.05 (high). We assume that $\rho = 0.99$ so a signal is very informative about next period's marginal costs and signaling incentives are strong.

Figure B.2: Equilibrium Prices in the Two-Type Marginal Cost Model (parameters described in the text)

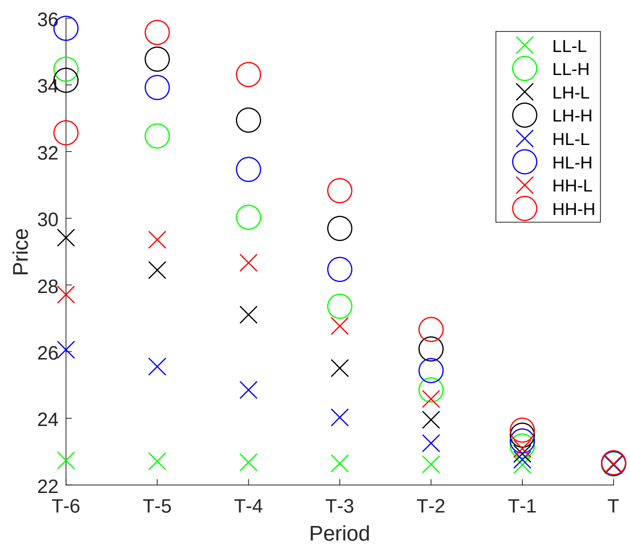


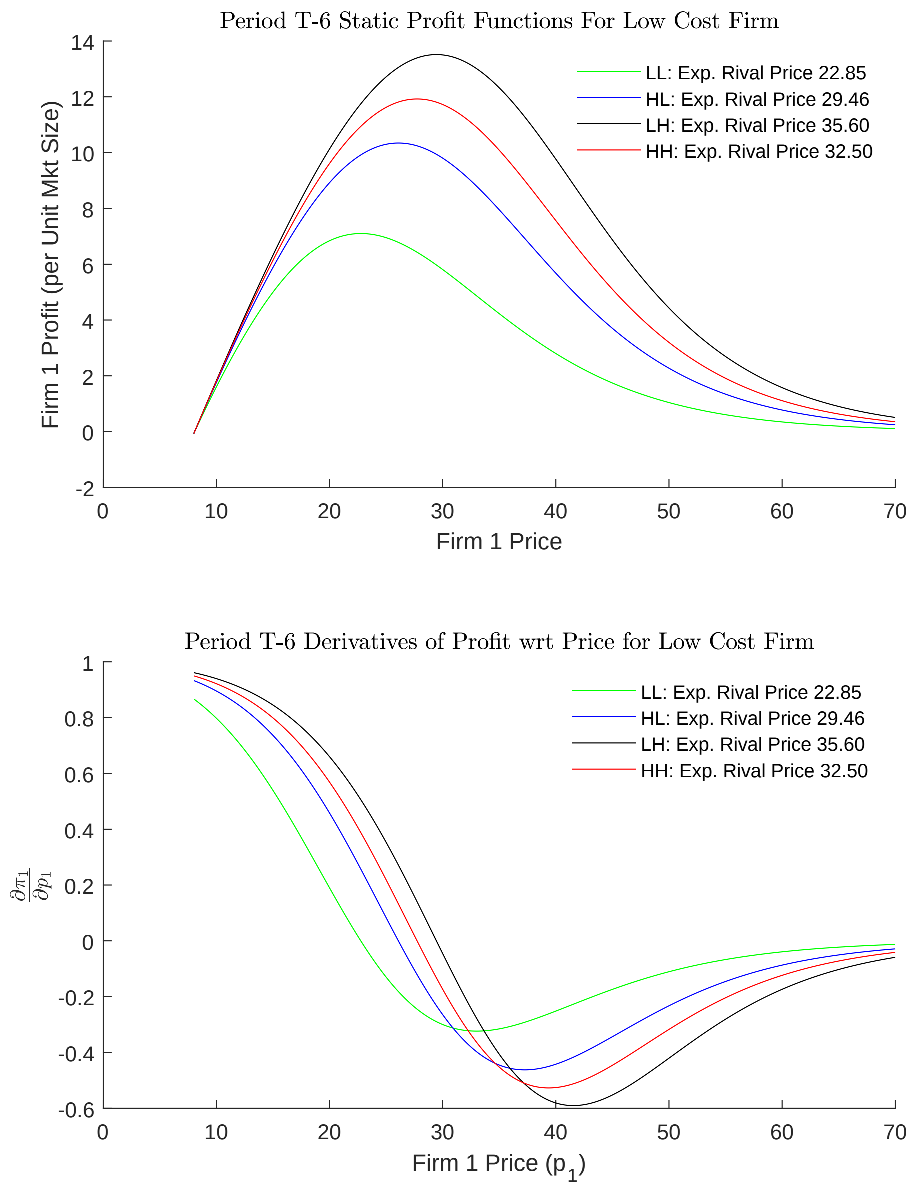
Figure B.2 shows the full set of eight equilibrium prices in each period as we move backwards from the end of the game. The legend denotes states by {"the firm's perceived cost in $t-1$", "its rival's perceived cost in $t-1$" - "the firm's realized marginal cost in t"} so blue indicates prices for a firm whose perceived marginal cost in the previous period was high, its rival's perceived previous period cost was low, and a cross (circle) indicates that the firm's current cost is low (high).

The green crosses (LL-L) remain almost unchanged across periods, as they represent static best responses when both players know that their rival is very likely to be setting the same price, but, as we move earlier in the game, the remaining prices increase, because they involve either signaling (by a $\bar{c}$ firm) or a static best response to a rival who is likely to be raising its price to signal.

In period $T-6$ the order of the prices changes with the HH-H price (red circle) below the

HL-H price (blue circle). This implies that in period $T-7$, a firm that believes its rival is likely to be high cost, is more likely to increase its rival's next period $(T-6)$ price if it (the firm) is believed to be *low* cost than if it is believed to be high cost. As profits increase in the rival's price, this will lead belief monotonicity to be violated.

Figure B.3: Period $T-6$ Profit Functions in the Two-Type Game



Why does the order of the red and blue circles switch? It reflects changes in both the incentive to signal (i.e., the possible effect on future prices) and the cost of signaling (i.e., the effect on current profits). Recall that in the two-type model the equilibrium price of the $\overline{c}$ type is determined by the lowest price that the low-cost firm would be unwilling to set even if choosing it would lead to it being perceived as high cost. Consider the cost, in terms of foregone period $T-6$ profit, for a low-cost firm of raising its price. The upper panel of Figure B.3 shows the period $T-6$ one-period profit functions for a low cost firm given different beliefs about previous firm types and the expected price of the rival.[49] The lower panel shows the corresponding derivatives of the profit function with respect to the firm's own price. For prices above \$34, the marginal loss in profit from a price increase is greater for a red firm (i.e., a firm likely to face a high cost rival) than a blue firm (i.e., a low cost rival) so it is less costly for the blue firm to raise its price.[50]

Now consider the incentive of a low-cost firm to signal (i.e., to pretend to be high-cost). The incentive of an HL (blue) firm to signal a high cost in period $T-6$ is that it is very likely to lead to its rival setting the black cross, rather than the green cross, price in period $T-5$. This difference is large, so that the incentive to signal is strong. The incentive of an HH (red) firm to signal is that this will very likely lead to it facing the red, rather than the blue, circle price in period $T-5$. These period $T-5$ prices are closer together (than the black and green crosses) so the incentive to signal will tend to be weaker. The cost and the incentive effects together lead to a reversal of the order of the period $T-6$ equilibrium prices, causing belief monotonicity to fail in period $T-7$.

[49] For example, an HL firm expects to face a low-cost LH firm (setting a black cross price) with probability 0.99, so the expected rival price is \$29.46.

[50] The crossing of the derivative functions reflects the failure of strategic complementarity (defined as $\frac{\partial^2 \pi_i}{\partial p_i \partial p_j} > 0$) for logit-based demand when prices are significantly above static profit-maximizing levels. The intuition is that, as a rival's price increases, the incentive for a firm to reduce its (high) price towards the static best response price can increase.

B.2 Alternative Sources of Asymmetric Information.

While it is plausible that, in many industries, firms have some private information about their marginal costs and that whatever is unobserved is likely to be serially correlated, our results are not dependent on assuming that it is marginal costs that are privately observed. In this Appendix we consider three examples where marginal costs are fixed and known and the asymmetric information is embedded in a different part of the profit function. In each case we show that equilibrium prices can be significantly higher, and more volatile, than in the CI or static incomplete information versions of the model. The fact that other formulations generate similar results is not surprising, but we perform the calculations in order to emphasize the point that we are not tied to the marginal cost assumption. In all cases, we assume single-product duopolists, as in Section 3, and we solve the continuous type, infinite horizon version of our model. The demand parameters also take on their baseline values from Section 3, and marginal cost of each firm is held fixed at 8.

Variant 1: Weights on Profits and Revenues. In the first variant, we allow for there to be uncertainty about the weight that each firm places on profits rather than revenues. A number of theoretical and empirical papers study whether managers want to maximize profits or alternative outcome variables, and whether shareholders might strategically choose to incentivize managers to deviate from profit maximization (e.g., Sklivas (1987), Katz (1991), Murphy (1999), De Angelis and Grinstein (2014)). The empirical literature suggests that managers are affected by a variety of incentives that may be complicated for outsiders to evaluate and which may vary over time, depending on oversight from shareholders or corporate boards, and financial constraints.

Without assuming a particular theory of governance, we suppose that the weight placed on profits by firm i in period t is $\tau_{i,t}$ and that this variable lies on the interval $[0.89, 0.9]$, with the remaining weight on firm revenues. As before, we suppose that the variable evolves according to a truncated AR(1) process, with $\rho = 0.8$. The standard deviation of the innovations is chosen so that, as for our baseline model where marginal costs are private information, the probability that a type will transition from the highest point of the support in one period to a value in the lower half of the support in the next period is 0.32.

The first panel of Table B.1 reports the average CI price when both firms (are known to)

Table B.1: Price Effects in Models Where Alternative Elements of Firm Objective Functions are Private Information

	Static BNE		Infinite Horizon Signaling Model	
Model	Mean Price	Std. Dev. Price	Mean Price	Std. Dev. Price
Weight on Firm Profits vs. Revenues is Private Information	21.79	0.01	24.45	0.59
Weight on Profits on [0.89,0.9]				
$\rho = 0.8$, std. dev. innovation 0.0044				
(eqm. prices if firms are known to maximize profits, i.e., if weight on profit equals 1, are 22.59)				
Weight on Firm vs. Industry Profits is Private Information	22.66	0.02	25.34	0.59
Weight on Industry Profit [0.00,0.02]				
$\rho = 0.8$, std. dev. innovation 0.0088				
(eqm. prices if firms are known to maximize own profits are 22.59)				
Size of Each Firm's Loyal Market is Private Information	25.17	0.07	27.56	0.59
Size of Loyal Market [0.10,0.12] (as fraction of duopoly market)				
$\rho = 0.8$, std. dev. innovation 0.0088				
(eqm. prices when loyal markets known to be 11% of duopoly markets are 25.17)				

maximize profits is 22.59. When a firm places some weight on revenues, it will tend to set a lower price, and the average static BNE or CI price when the profit weight lies on $[0.89, 0.90]$ is 21.79. However, with signaling, average prices increase significantly: in this example, the average Markov Perfect Bayesian Equilibrium price is 8.2% above the average price level *when both firms are known to maximize profits*, with profits increasing by 18%. This example suggests there may be some advantage to shareholders if they keep managers' incentives opaque to rivals even in markets where firms set prices for differentiated products.[51]

Variant 2: Weight on Profits of Other Firms in the Industry. In the empirical Industrial Organization literature, it is common to model tacitly collusive behavior in a reduced-form way by generalizing static first-order conditions to allow for each firm to place some weight on the profits of other firms in the same market (Porter (1983), Bresnahan (1989), Miller and Weinberg (2017)). This type of formulation could also be rationalized by models where participants in financial markets become more optimistic about a firm's prospects when its rivals announce high profits (Rotemberg and Scharfstein (1990)) or by models where firms maximize the overall returns of shareholders who hold stock in competitors (O'Brien and Salop (1999), Azar, Schmalz, and Tecu (2018)).

We consider a model where rivals have some limited uncertainty about the weight that a firm places on its own profit rather than the profit of the industry. Specifically we assume that each firm places a weight $\tau_{i,t}$ of $[0.98, 1]$ on its own profits, and $1 - \tau_{i,t}$ on the profits of the industry as a whole (of course, its own profits also contribute to industry profits). We assume that the transition process has ρ=0.8 and $\sigma = 0.0088$, which means that the probability of a type transitioning from the highest point of the support to below the median is 0.32, as in the first example. As can be seen in the second panel of Table B.1, the effect is, once again, to raise prices substantially in the dynamic game with asymmetric information.

Variant 3: Demand Shocks. Our experience in seminars is that many economists believe it is more intuitive that some aspect of demand will be private information to the firm than marginal costs will be.

Some formulations of demand uncertainty give rise to signaling incentives that would be

[51] The usual explanation for why shareholders might want to commit to incentivizing their managers to place some weight on revenues comes from quantity-setting models where other firms will reduce their output when a firm's managers are committed to increase their output. In our model it is uncertainty about what firms are trying to maximize that causes equilibrium prices to rise, through the mechanism of signaling.

qualitatively different from the ones in our framework. For example, suppose that demand has a logit structure and that each firm has private information about the serially correlated and unobserved quality of its product. Duopolist firms observe each other's prices but not quantities, so that prices are informative about quality. A firm with higher quality will want to charge a higher price, but its rival's optimal price will likely decrease in the firm's quality, so it is unclear whether a firm will want to be perceived as high quality or as low quality. This is likely to be a case where only some type of pooling equilibrium exists.

Here we consider a simple example where firms do have incentives to raise prices to signal that their demand is high. Suppose that each firm sells its products in two markets. In one market, the firms compete as duopolists, but in the other market the firm is a monopolist (so for example, both firms are in market A, firm 1 is the only firm in market B, and firm 2 is the only firm in market C). Due to the possibility of arbitrage, or some other constraint, each firm can only set one price across the markets. One rationalization of this setup would be that each firm has some loyal or locked-in customers, but that additional consumers are competed for. Product quality is known, but firms are uncertain about the size of their rival's loyal market. Normalizing the size of the common market to 1, the sizes of the loyal markets lie between $[0.1, 0.12]$. The utility specification is the same as before except loyal customers only choose between a single product and the outside good. The transition assumptions are the same as in variant 2. In this formulation, firms will set prices based on the weighted average marginal revenues from the two markets, and when the size of their monopoly market is larger they will prefer higher prices. A firm will therefore have incentive to raise its price to signal that its monopoly market is larger.

The results are presented in the third panel of Table B.1. The addition of the loyal market, where a firm's demand is less elastic, raises prices under all information structures, but the average signaling equilibrium prices are 10% higher than the prices under CI or in a static game with asymmetric information.

C Existence and Uniqueness of a Fully Separating Equilibrium in a Finite Horizon Game with Linear Demand

As discussed in the text, Mailath (1989) and Mester (1992) provide proofs of the existence and uniqueness of a fully separating equilibrium in a two-period duopoly, linear demand, continuous cost price-setting game and a three-period duopoly, linear demand, continuous cost quantity-setting games respectively. This Appendix presents a theoretical proof of existence and uniqueness of a fully-separating Markov Perfect Bayesian Equilibrium for a finite-horizon duopoly pricing game with linear demand and marginal costs that are private information, under a condition that the range of costs is "small enough" so that the single-crossing condition holds. As explained in the text, we have to rely on computational analysis when assuming nonlinear demand or an infinite horizon, and in our application we assume both.[52] However, we include our proof for the linear demand and finite horizon case for completeness.

We make the following specific assumptions on the model. There are two firms, and i will index the firm.

Assumptions

A1 (linear demand). $q_{i,t} = a_i - b_{1,i}p_{i,t} + b_{2,i}p_{j,t}$, $b_{1,i} > b_{2,i} > 0$.

A2 (positive demand). The intercepts a are large enough that for all of the prices charged on the equilibrium path, both firms will have positive output.

A3 (continuous cost interval). The marginal costs of each firm, $c_{i,t}$, lie on compact intervals where $[\underline{c_i}, \overline{c_i}]$ where $\overline{c_i} > \underline{c_i} > 0$.

A4 (cost transitions). Costs evolve independently according to first-order Markov processes with conditional densities $\Psi_i(c_{i,t}|c_{i,t-1})$, where the conditional density functions are smooth in $c_{i,t}$ and $c_{i,t-1}$ and strictly positive for all $[\underline{c_i}, \overline{c_i}]$. $E(c_{i,t}|c_{i,t-1})$ is continuous and strictly increasing in $c_{i,t-1}$.

A5 (discount factor). There is a common discount factor $0 < \beta < 1$.

The statement of the results and the proof will use the following notation.

- $\pi_{i,t}$ denotes per-period profits in period t. $\pi_{i,t} = (p_{i,t} - c_{i,t})q_{i,t}(p_{i,t}, p_{-i,t})$.

[52] Our proof is for two firms that may be asymmetric. Extending the proof to three symmetric firms is straightforward.

- $V_{i,t}(c_{i,t-1}, \widehat{c_{i,t-1}}, c_{j,t-1})$ is i's value at the beginning of period t, before $c_{i,t}$ is revealed, when it is perceived to have cost $\widehat{c_{i,t-1}}$, and its real cost is $c_{i,t-1}$, and it believes that j's $t-1$ cost was $c_{j,t-1}$.

- $\Pi^{i,t}(c_{i,t}, \widehat{c_{i,t}}, p_{i,t}, c_{j,t-1})$ ("signaling payoff function") represents the expected current and future profits (given equilibrium behavior in future periods) of firm i in period t, when it sets price $p_{i,t}$, has cost $c_{i,t}$ and is perceived, at the end of the period, as having cost $\widehat{c_{i,t}}$. $c_{j,t-1}$ is i's perception of j's cost in period $t-1$. In equilibrium, this perception will be correct so we denote it simply by $c_{j,t-1}$. $\Pi^{i,t}(c_{i,t}, \widehat{c_{i,t}}, p_{i,t}, c_{j,t-1})$ is implicitly conditioned on j's period t pricing strategy, which will involve j setting a price with an average of $\overline{p_{j,t}}$ and which i assumes will reveal $c_{j,t}$. $\Pi^{i,t}_k(c_{i,t}, \widehat{c_{i,t}}, p_{i,t}, c_{j,t-1})$ denotes the derivative of $\Pi^{i,t}(c_{i,t}, \widehat{c_{i,t}}, p_{i,t}, c_{j,t-1})$ with respect to the k^{th} argument.

- Prices (the proof will indicate conditioning arguments where necesary):

  - $p^*_{i,t}$ is i's equilibrium strategy in a fully separating MBPE (i.e., it is a function);
  - $p^{BR}_{i,t}$ is i's separating best response pricing function given some separating strategy (not necessarily the equilibrium strategy) by j;
  - $p^{**}_{i,t}$ is a price that is a statically optimal best response (i.e., maximizes i's current profits) given j's strategy;
  - $\overline{p_{j,t}}$ is the average price set by j when it uses a particular strategy; and,
  - our description of separating pricing strategies will refer to "initial values", which will reflect a $p^{**}_{i,t}$ price determined as the solution to a static profit maximization problem when $c_{i,t} = \underline{c_i}$, and, the "increment" which refers to the additional price above this initial value that may reflect signaling behavior.

C.1 Preliminary Results.

We begin with a useful Lemma.

Lemma 1 *In a fully separating Markov Perfect Bayesian Equilibrium, play on the equilibrium path will have the following properties, (L-i) $p^*_{i,t}$ will be a function of $c_{i,t}$ and the costs $c_{i,t-1}$ and $c_{j,t-1}$ revealed by prices at $t-1$; (L-ii) the only effect of $c_{i,t-1}$ on $p^*_{i,t}$ is through the effect that it* *will have on the expected value of $p_{j,t}$; (L-iii) i's period t price, and the inference that j makes about $c_{i,t}$, based on this price, will affect i's profits in t and $t+1$ only.*

Proof. (L-i) In a fully separating equilibrium, prices at $t-1$ will reveal marginal costs at $t-1$ and the first-order Markovian assumption on the Ψ_is implies that costs at $t-1$ contain all available information from earlier periods about costs. The Markovian equilibrium assumption implies that strategies depend on payoff-relevant state variables (current costs) and beliefs about those variables, only. This implies that strategies can be functions of $c_{i,t}$ (which is private information to i when $p_{i,t}$ is chosen), $c_{i,t-1}$ and $c_{j,t-1}$ only.

(L-ii) The equilibrium choice of $p^*_{i,t}$ will depend on its effect on expected profits in future periods and expected profits at t. Property (L-i) implies that given $p_{i,t}$, which reveals $c_{i,t}$, $c_{i,t-1}$ will not affect what happens at $t+1$. Expected profits in period t are $(p_{i,t}-c_{i,t})(a_i - b_{1,i}p_{i,t} + b_{2,i}\overline{p_{j,t}})$ so $c_{i,t-1}$ can only affect i's payoffs through its effect on $\overline{p_{j,t}}$.

(L-iii) Suppose that instead of equilibrium price $p^*_{i,t}$, i sets a price $p'_{i,t}$ in the range of the equilibrium price function. $t+1$ strategies specify an optimal strategy for i given $c_{i,t+1}$, $c_{j,t}$ and the cost implied by $p'_{i,t}$, and it will be optimal to use these strategies at $t+1$ (because of property (L-ii)), so $t+1$ strategies will correctly reveal $c_{i,t+1}$. Therefore charging $p'_{i,t}$ not $p^{**}_{i,t}$ only affects profits at t and $t+1$. ■

Our results characterizing firm i's separating best response function in period t, given a fully revealing pricing strategy, of any form, by j and the assumed form of strategies at $t+1$, are based on the following theorems which are adapted from Mailath (1987).

Theorem 1 *Adapted from Theorems 1 and 2, and the Corollary, in Mailath (1987). If (MT-i) $\Pi^{i,t}(c_{i,t}, \widehat{c_{i,t}}, p_{i,t}, c_{j,t-1})$ is smooth in arguments $(c_{i,t}, \widehat{c_{i,t}})$, (MT-ii) $\Pi^{i,t}_2(c_{i,t}, \widehat{c_{i,t}}, p_{i,t}, c_{j,t-1}) > 0$ [belief monotonicity], (MT-iii) $\Pi^{i,t}_{13}(c_{i,t}, \widehat{c_{i,t}}, p_{i,t}, c_{j,t-1}) > 0$ [type monotonicity], (MT-iv) $\Pi^{i,t}_3(c_{i,t}, \widehat{c_{i,t}}, p_{i,t}, c_{j,t-1}) = 0$ for only one p_i, and for this p_i, $\Pi^{i,t}_{33}(c_{i,t}, \widehat{c_{i,t}}, p_{i,t}, c_{j,t-1}) < 0$ [strict quasi-concavity], (MT-v) there exists $k > 0$ such that $\Pi^{i,t}_{33}(c_{i,t}, \widehat{c_{i,t}}, p_{i,t}, c_{j,t-1}) \geq 0$ implies $\left|\Pi^{i,t}_3(c_{i,t}, \widehat{c_{i,t}}, p_{i,t}, c_{j,t-1})\right| > k$, then a pricing function $p^{BR}_{i,t}(c_{i,t}, c_{j,t-1})$ that solves the differential equation*

$$\frac{\partial p^{BR}_{i,t}(c_{i,t}, c_{j,t-1})}{\partial c_{i,t}} = -\frac{\Pi^{i,t}_2(c_{i,t}, c_{i,t}, p_{i,t}, c_{j,t-1})}{\Pi^{i,t}_3(c_{i,t}, c_{i,t}, p_{i,t}, c_{j,t-1})}$$

and has a lower initial value condition where $p^{BR}_{i,t}(\underline{c_i}, c_{j,t-1})$ solves $\Pi^{i,t}_3(\underline{c_i}, \underline{c_i}, p^{BR}_{i,t}(\underline{c_i}, c_{j,t-1}), c_{j,t-1}) =$

0 *is the unique fully separating best response function if a fully separating best response exists.*

Theorem 2 *Adapted from Theorem 3 in Mailath (1987). Suppose assumptions (MT-i)-(MT-v) in Theorem 1 hold. If (MT-vi), for* $(\widehat{c_i}, p)$ *in the graph of* $p_{i,t}^{BR}(c_{i,t}, c_{j,t-1})$, $\frac{\Pi_3^{i,t}(c_{i,t},\widehat{c_{i,t}},p_{i,t},c_{j,t-1})}{\Pi_2^{i,t}(c_{i,t},\widehat{c_{i,t}},p_{i,t},c_{j,t-1})}$ *is either strictly increasing or decreasing in* $c_{i,t}$ *[single-crossing], then the fully separating best response described in Theorem 1 exists.*

C.2 Main Result.

The following theorem gives our main result.

Theorem 3 *If* $\overline{c_i} - \underline{c_i}$ *is small enough for all* i*, in any finite horizon game there will exist a unique fully separating MPBE where, on the equilibrium path, firm* i*'s equilibrium pricing strategy* $p_{i,t}^*(c_{i,t}, c_{i,t-1}, c_{j,t-1})$ *in any period* $t < T$ *has the form of the best response function described in Theorem 1. In period* T *firms will choose static payoff-maximizing prices given their beliefs about rivals' costs in period* $T-1$*. In periods* $t < T$*, pricing strategies will have the following features: (T-i) (a) the initial values (i.e., static best response prices when* $c_{i,t} = \underline{c_i}$*) are functions of* $c_{j,t-1}$ *and* $c_{i,t-1}$ *only (in the following we will denote the function that determines the initial value* $g_{i,t}(c_{j,t-1}, c_{i,t-1})$*), and (b) the increment above the initial value (a function* $f_{i,t}(c_{i,t}, c_{j,t-1})$*) is a continuous function of* $c_{i,t}$ *and* $c_{j,t-1}$ *only, and in particular it does not depend on* $\overline{p_{j,t}}$*; (T-ii) for all* $c_{i,t} > \underline{c_i}$ *the price charged is always above the static best response price for* $c_{i,t}$*, (T-iii) the effect of* $c_{j,t-1}$ *on the increment only comes through its effect on* i*'s belief about the distribution of* $c_{j,t+1}$*, and (T-iv) (a)* i*'s pricing function is continuous and strictly increasing in* $c_{i,t}$*, (b)* i*'s pricing function is continuous and strictly increasing in* $\overline{p_{j,t}}$*, (c)* i*'s pricing function is continuous and strictly increasing in* $c_{i,t-1}$ *and (*i*'s perception of)* $c_{j,t-1}$*.*

C.2.1 Proof.

The proof uses induction, showing that if strategies have this form in periods $t+1$,...,$T-1$ there will exist a unique MPBE with the required form in any period $t < T-1$. We then show that the form of equilibrium strategies in period T will lead to strategies that have the specified form in period $T-1$.

Period $t < T-1$. The logic of the proof for period t is to show that the conditions required for Mailath's theorems hold given Lemma 1 and the assumed equilibrium form of pricing behavior in $t+1$. This shows that there will be a unique best response pricing function for each firm given any separating strategy of the other firm. This will let us show some of the features specified above. We then show that there can be only one pair of pricing functions with these properties that are best responses to each other, and this will allow us to show the remaining features.

Uniqueness, Existence and Form of i's Fully Separating Best Response Function Given j's Strategy.

We go through the conditions required for Mailath's results in turn.

Condition (MT-i): $\Pi^{i,t}(c_{i,t}, \widehat{c_{i,t}}, p_{i,t}, c_{j,t-1})$ is smooth in arguments $(c_{i,t}, \widehat{c_{i,t}})$. Lemma 1 implies that

$$\Pi^{i,t}(c_{i,t}, \widehat{c_{i,t}}, p_{i,t}, c_{j,t-1}) = E\pi_{i,t}(c_{i,t}, p_{i,t}, \overline{p_{j,t}}) + \beta E(V_{i,t+1}(c_{i,t}, \widehat{c_{i,t}}, c_{j,t}|c_{j,t-1}))$$

where the second expectation is over the cost that j reveals in period t. $E\pi_{i,t}(c_{i,t}, p_{i,t}, \overline{p_{j,t}}) = (p_{i,t} - c_{i,t})(a_i - b_{1,i}p_{i,t} + b_{2,i}\overline{p_{j,t}})$ which is smooth in $c_{i,t}$. Profits in $t+1$ will be equal to $(p_{i,t+1} - c_{i,t+1})(a_i - b_{1,i}p_{i,t+1} + b_{2,i}p_{j,t+1})$ and smoothness of the period-t expectation of these profits follows from the assumed smoothness of the Ψ_i conditional densities (A4) and the continuity of the pricing functions (T-i/T-iv). Similar logic (and the results concerning period T prices below) implies that the period-t expectation of discounted profits in $t+2$ and future periods will also be continuous in $c_{i,t}, c_{j,t-1}$ and $\widehat{c_{i,t}}$. Therefore $\beta E(V_{i,t+1}(c_{i,t}, \widehat{c_{i,t}}, c_{j,t}|c_{j,t-1}))$ will be smooth in $c_{i,t}, \widehat{c_{i,t}}$ and $c_{j,t-1}$.

Condition (MT-ii): $\Pi_2^{i,t}(c_{i,t}, \widehat{c_{i,t}}, p_{i,t}, c_{j,t-1}) > 0$. From Lemma 1, $\widehat{c_{i,t}}$ only affects future profits in period $t+1$ given equilibrium play from $t+1$ forwards (L-iii). Denote expected profits in period $t+1$ when j charges an expected price $\overline{p_{j,t+1}}(\widehat{c_{i,t}}, c_{j,t})$, $E\pi_{i,t+1}(c_{i,t+1}, p_{i,t+1}, \overline{p_{j,t+1}}(\widehat{c_{i,t}}, c_{j,t}))$, so

$$\Pi_2^{i,t}(c_{i,t},\widehat{c_{i,t}},p_{i,t},c_{j,t-1}) = \int\int \frac{\partial E\pi_{i,t+1}(c_{i,t+1},p_{i,t+1},\overline{p_{j,t+1}}(\widehat{c_{i,t}},c_{j,t}))}{\partial\widehat{c_{i,t}}}\Psi_i(c_{i,t+1}|c_{i,t})\Psi_j(c_{j,t}|c_{j,t-1})dc_{i,t+1}dc_{j,t}$$

Given that $\frac{\partial\overline{p_{j,t+1}}(\widehat{c_{i,t}},c_{j,t})}{\partial\widehat{c_{i,t}}} > 0$ (T-iv (c)), it is sufficient to show that $\frac{\partial E\pi_{i,t+1}(c_{i,t+1},p_{i,t+1},\overline{p_{j,t+1}})}{\partial\overline{p_{j,t+1}}} > 0$.

Express the price that i charges in $t+1$ as $p_{i,t+1} = p^{**}_{i,t+1}(\overline{p_{j,t+1}},c_{i,t+1})+p'$, where $p^{**}_{i,t+1}(\overline{p_{j,t+1}},c_{i,t+1})$ is the static profit-maximizing best response to $\overline{p_{j,t+1}}$ given $c_{i,t+1}$ and $p' \geq 0$ is an increment above the static best response price. Linear demand implies that i's expected $t+1$ profit is

$$\begin{aligned}
E\pi'_{i,t+1}(c_{i,t+1},p',\overline{p_{j,t+1}}) &= (p^{**}_{i,t+1}(\overline{p_{j,t+1}},c_{i,t+1})+p'-c_{i,t+1})(a_i - b_{1,i}(p^{**}_{i,t+1}(\overline{p_{j,t+1}},c_{i,t+1})+p')+b_{2,i}\overline{p_{j,t+1}})\\
&= E\pi'_{i,t+1}(c_{i,t+1},0,\overline{p_{j,t+1}}) + \int_0^{p'} \frac{\partial E\pi'_{i,t+1}(c_{i,t+1},x,\overline{p_{j,t+1}})}{\partial x}dx\\
&= E\pi'_{i,t+1}(c_{i,t+1},0,\overline{p_{j,t+1}}) + \int_0^{p'} (-2b_{i,1}x)dx
\end{aligned}$$

where the last line uses the facts that

$$\frac{\partial E\pi'_{i,t+1}(c_{i,t+1},x,\overline{p_{j,t+1}})}{\partial x} = a_i - 2b_{1,i}p^{**}_{i,t+1}(\overline{p_{j,t+1}},c_{i,t+1}) - 2b_{1,i}x + b_{2,i}\overline{p_{j,t+1}} + b_{1,i}c_{i,t+1},$$

and

$$a_i - 2b_{1,i}p^{**}_{i,t+1}(\overline{p_{j,t+1}},c_{i,t+1}) + b_{2,i}\overline{p_{j,t+1}} + b_{1,i}c_{i,t+1} = 0,$$

as $p^{**}_{i,t+1}(\overline{p_{j,t+1}},c_{i,t+1})$ is the static profit-maximizing price, so that $\frac{\partial E\pi'_{i,t+1}(c_{i,t+1},x,\overline{p_{j,t+1}})}{\partial x} = -2b_{1,i}x$.

Therefore,

$$\begin{aligned}
\frac{\partial E\pi'_{i,t+1}(c_{i,t+1},p',\overline{p_{j,t+1}})}{\partial\overline{p_{j,t+1}}} &= \frac{\partial E\pi'_{i,t+1}(c_{i,t+1},0,\overline{p_{j,t+1}})}{\partial\overline{p_{j,t+1}}}\\
&= b_{2,i}(p^{**}_{i,t+1}(\overline{p_{j,t+1}},c_{i,t+1}) - c_{i,t+1}) > 0.
\end{aligned}$$

where the final step uses the envelope-theorem as $p^{**}_{i,t+1}(\overline{p_{j,t+1}},c_{i,t+1})$ is the static profit-maximizing price.

Condition (MT-iii): $\Pi_{13}^{i,t}(c_{i,t}, \widehat{c_{i,t}}, p_{i,t}, c_{j,t-1}) > 0$.

$$\frac{\partial \Pi^{i,t}(c_{i,t}, \widehat{c_{i,t}}, p_{i,t}, c_{j,t-1})}{\partial p_{i,t}} = a_i - 2b_{i,1}p_{i,t} + b_{2,i}\overline{p_{j,t}} + b_{i,1}c_{i,t}$$

as, conditional on $\widehat{c_{i,t}}$, $p_{i,t}$ only affects period t profits. Therefore,

$$\Pi_{13}^{i,t}(c_{i,t}, \widehat{c_{i,t}}, p_{i,t}, c_{j,t-1}) = \frac{\partial \Pi^{i,t}(c_{i,t}, \widehat{c_{i,t}}, p_{i,t}, c_{j,t-1})}{\partial p_{i,t} \partial c_{i,t}} = b_{1,i} > 0$$

Condition (MT-iv): $\Pi_{3}^{i,t}(c_{i,t}, \widehat{c_{i,t}}, p_{i,t}, c_{j,t-1}) = 0$ for only one $p_{i,t}$, and for this $p_{i,t}$, $\Pi_{33}^{i,t}(c_{i,t}, \widehat{c_{i,t}}, p_{i,t}, c_{j,t-1})$ < 0.

$$\Pi_{33}^{i,t}(c_{i,t}, \widehat{c_{i,t}}, p_{i,t}, c_{j,t-1}) = -2b_{1,i} < 0 \ \forall p_{i,t}$$

so $\Pi^{i,t}(c_{i,t}, \widehat{c_{i,t}}, p_{i,t}, c_{j,t-1})$ will have a unique maximum in $p_{i,t}$.

Condition (MT-v): there exists $k > 0$ such that if $\Pi_{33}^{i,t}(c_i, \widehat{c}_i, p_i, c_{j,t-1}) \geq 0$ then $\left|\Pi_{3}^{i,t}(c_i, \widehat{c}_i, p_i, c_{j,t-1})\right| >$ k. As $\Pi_{33}^{i,t}(c_i, \widehat{c}_i, p_i, c_{j,t-1}) < 0$ for all $p_{i,t}$, the condition is trivially satisfied.

Therefore, based on Theorem 3, if a fully separating best response function in period t exists, it is uniquely characterized as $p_{i,t}^{BR}(c_{i,t}, c_{j,t-1})$ as the solution to a differential equation

$$\frac{\partial p_{i,t}^{BR}(c_{i,t}, c_{j,t-1})}{\partial c_{i,t}} = -\frac{\Pi_{2}^{i,t}(c_{i,t}, c_{i,t}, p_{i,t}, c_{j,t-1})}{\Pi_{3}^{i,t}(c_{i,t}, c_{i,t}, p_{i,t}, c_{j,t-1})}$$

with a lower initial condition price $p_{i,t}^{BR}(\underline{c_i}, c_{j,t-1})$ that solves $\Pi_{3}^{i,t}(\underline{c_i}, \underline{c_i}, p_{i,t}^{BR}(\underline{c_i}, c_{j,t-1}), c_{j,t-1}) = 0$.

Period t Pricing Function Properties, Part I

Before discussing single-crossing, we can now prove some features of period-t pricing functions given this characterization of best responses.

Feature (T-ii): the price charged is always above the static best response price for all $c_{i,t} > \underline{c_i}$.

Proof: as $\Pi_2^{i,t}(c_{i,t}, c_{i,t}, p_{i,t}, c_{j,t-1}) > 0$, and is independent of the value of $p_{i,t}$, and $\Pi_3^{i,t}(c_{i,t}, c_{i,t}, p_{i,t}, c_{j,t-1}) < 0$ for prices above the static best response price, and $\Pi_3^{i,t}(c_{i,t}, c_{i,t}, p_{i,t}, c_{j,t}) \to 0$ as $p_{i,t}$ approaches the static best response price for any $c_{i,t}$, the solution to the differential equation for a specific $c_{i,t}$ will be greater than the static best response price given $c_{i,t}$ except at $\underline{c_i}$.

Feature (T-i(b)): the increment above the initial value is a function of $c_{i,t}$ and $c_{j,t-1}$ only, and it does not depend on $\overline{p_{j,t}}$.

Proof: the initial value solves $\Pi_3^{i,t}(\underline{c_i}, \underline{c_i}, p^*_{i,t}(\underline{c_i}), c_{j,t-1}) = 0$, i.e., it is a static best response when $c_{i,t} = \underline{c_i}$ to the expected price $\overline{p_{j,t}}$. As the numerator in the differential equation, $\Pi_2^{i,t}(c_{i,t}, \widehat{c_{i,t}}, p_{i,t}, c_{j,t-1})$, is independent of $p_{i,t}$ and $\Pi_3^{i,t}(c_{i,t}, \widehat{c_{i,t}}, p_{i,t}, c_{j,t-1})$ depends only on the increment of $p_{i,t}$ above the intercept, the increment depends only on $c_{i,t}$ and (possibly) $c_{j,t-1}$.[53]

Feature (T-iii): the effect of $c_{j,t-1}$ on the increment only comes through its effect on i's belief about the distribution of $c_{j,t+1}$.

Proof: $c_{j,t-1}$ affects $\overline{p_{j,t}}$ and i's period t belief about the distribution of $c_{j,t+1}$, which will affect i's expectation of $\overline{p_{j,t+1}}$. Given T-i(b), $\overline{p_{j,t}}$ does not affect the increment. From Lemma 1 (L-ii), at the start of period $t+1$, the expectation of $\overline{p_{j,t+1}}$ will depend only on $c_{j,t}$ (revealed by j's period t price) and $\widehat{c_{i,t}}$. Therefore the only effect that $c_{j,t-1}$ can have on the period t increment, which is set before $p_{j,t}$ is revealed, is that it affects i's beliefs about the distribution of $c_{j,t+1}$.

Feature (T-iv): (a) the pricing function is increasing and continuous in $c_{i,t}$.

[53] The proof of (MT-ii) shows that $\Pi_3^{i,t}$ only depends on the increment of $p_{i,t}$ above the static best response price for $c_{i,t}$ (not the initial value which is the best response for $\underline{c_i}$). However, given linear demand, static best responses are given by

$$p^{**}_{i,t} = \frac{a_i}{2b_{1,i}} + \frac{c_{i,t}}{2} + \frac{b_{2,i}}{2b_{1,i}}\overline{p_{j,t}}$$

so the increment of the static best response price above the static best response for $c_{i,t} = \underline{c_i}$ only depends on $c_{i,t} - \underline{c_i}$.

Proof: (a) as $\Pi_2^{i,t}(c_{i,t}, c_{i,t}, p_{i,t}, c_{j,t-1}) > 0$ and $\Pi_3^{i,t}(c_{i,t}, c_{i,t}, p_{i,t}, c_{j,t-1}) < 0$ above the static best response price, the pricing function must be increasing in $c_{i,t}$.

Single-Crossing.

Condition (MT-vi): we need to show that, in the graph of $(\widehat{c_{i,t}}, p_{i,t})$, $\frac{\Pi_3^{i,t}(c_{i,t},\widehat{c_{i,t}},p_{i,t},c_{j,t-1})}{\Pi_2^{i,t}(c_{i,t},\widehat{c_{i,t}},p_{i,t},c_{j,t-1})}$ is either strictly increasing or decreasing in $c_{i,t}$. This amounts to showing that $\frac{\partial \frac{\Pi_3^{i,t}(c_{i,t},\widehat{c_{i,t}},p_{i,t},c_{j,t-1})}{\Pi_2^{i,t}(c_{i,t},\widehat{c_{i,t}},p_{i,t},c_{j,t-1})}}{\partial c_{i,t}}$ is either positive or negative within the graph of $(\widehat{c_{i,t}}, p_{i,t})$

$$\frac{\partial \frac{\Pi_3^{i,t}(c_{i,t},\widehat{c_{i,t}},p_{i,t},c_{j,t-1})}{\Pi_2^{i,t}(c_{i,t},\widehat{c_{i,t}},p_{i,t},c_{j,t-1})}}{\partial c_{i,t}} = \frac{\Pi_2^{i,t}(c_{i,t},\widehat{c_{i,t}},p_{i,t},c_{j,t-1})\frac{\partial \Pi_3^{i,t}(c_{i,t},\widehat{c_{i,t}},p_{i,t},c_{j,t-1})}{\partial c_{i,t}} - \Pi_3^{i,t}(c_{i,t},\widehat{c_{i,t}},p_{i,t},c_{j,t-1})\frac{\partial \Pi_2^{i,t}(c_{i,t},\widehat{c_{i,t}},p_{i,t},c_{j,t-1})}{\partial c_{i,t}}}{\left(\Pi_2^{i,t}(c_{i,t},\widehat{c_{i,t}},p_{i,t},c_{j,t-1})\right)^2}$$

The denominator is positive. As $\frac{\partial \Pi_3^{i,t}(c_{i,t},\widehat{c_{i,t}},p_{i,t},c_{j,t-1})}{\partial c_i} = b_1 > 0$, and $\Pi_2^{i,t}(c_{i,t},\widehat{c_{i,t}},p_{i,t}) > 0$ the first term in the numerator is strictly positive, and does not depend on $p_{i,t}$. Recognizing that $\frac{\partial \overline{p_{j,t+1}}}{\partial c_{i,t}\partial \widehat{c_{i,t}}} = 0$,

$$\frac{\partial \Pi_2^{i,t}(c_{i,t},\widehat{c_{i,t}},p_{i,t},c_{j,t-1})}{\partial c_{i,t}} = \beta b_{2,i}\int_{\underline{c_j}}^{\overline{c_j}}\int_{\underline{c_i}}^{\overline{c_i}} (p_{i,t+1}^{**}(c_{i,t+1},\overline{p_{j,t+1}}) - c_{i,t+1})\frac{\partial \overline{p_{j,t+1}}}{\partial \widehat{c_{i,t}}}\frac{\partial \Psi_i(c_{i,t+1}|c_{i,t})}{\partial c_{i,t}}\Psi_j(c_{j,t}|c_{j,t-1})dc_{i,t+1}dc_{j,t}.$$

$\frac{\partial \overline{p_{j,t+1}}}{\partial \widehat{c_{i,t}}}$ is positive (T-iv). With linear demand, the static mark-up, $p_{i,t+1}^{**}(c_{i,t+1},\overline{p_{j,t+1}}) - c_{i,t+1}$, will decrease in $c_{i,t+1}$, and given the assumptions on the densities Ψ_i, $\int (p_{i,t+1}^{**}(c_{i,t+1},\overline{p_{j,t+1}}) - c_{i,t+1})\frac{\partial \Psi_i(c_{i,t+1}|c_{i,t})}{\partial c_{i,t}}dc_{i,t+1} < 0$, but it will be bounded.

For prices at or above the static best response price, $\Pi_3^{i,t}(c_{i,t},\widehat{c_{i,t}},p_{i,t},c_{j,t-1}) \leq 0$, but, critically, $\Pi_3^{i,t}(c_{i,t},\widehat{c_{i,t}},p_{i,t},c_{j,t-1})$ must be close to 0 when $p_{i,t}$ is not too far above the static best response price. As the signaling price function is continuous and increasing in $c_{i,t}$, and is equal to the static best response price when $c_{i,t} = \underline{c_i}$, it follows that $\frac{\partial \frac{\Pi_3^{i,t}(c_i,\widehat{c_i},p_i)}{\Pi_2^{i,t}(c_i,\widehat{c_i},p_i)}}{\partial c_i} > 0$ when the interval $[\underline{c_i},\overline{c_i}]$ is small enough.

Therefore, from Theorem 2, the unique fully separating best response function described above exists.

A Unique MPBE in Period t Given the Form of the Best Response Functions.

The proof so far has chosen that, given a separating pricing strategy of j, i will have a unique fully separating best response that takes the required form. We now show that, with linear demand, the pair of separating functions used by i and j, given a pair $c_{i,t-1}$ and $c_{j,t-1}$, as best responses to each other, will be unique (i.e., there cannot be more than one distinct pair of best response functions that are best responses to each other).

Recall that the only effect of a change in $\overline{p_{j,t}}$ is on the intercept of i's pricing function. Therefore, holding fixed strategies in future periods, a change in j's period t strategy only translates i's best response pricing function upwards and downwards. It follows that there can only be a unique equilibrium if, for both i and j, $0 < \frac{\partial p^*_{i,t}}{\partial \overline{p_{j,t}}} < 1$.
Proof: $\frac{dp^*_{i,t}(c_i)}{d\overline{p_{j,t}}} = \frac{b_{2,i}}{2b_{1,i}}$, which, given A1, is strictly greater than zero and strictly less than one, as required.

Period t Pricing Function Properties, Part II.

We can now show the remaining features of the equilibrium pricing functions.

Feature (T-i(a)): the initial values (i.e., static best response prices when $c_{i,t} = \underline{c_i}$) are continuous functions of $c_{j,t-1}$ and $c_{i,t-1}$ only.

Proof: this follows directly from the Markovian assumption as $c_{j,t-1}$ and $c_{i,t-1}$ are sufficient to determine both players' beliefs about period t costs, and, given Theorem 3, to uniquely determine $\overline{p_{j,t}}$.

In the following, we will denote the function that determines the initial value $g_{i,t}(c_{j,t-1}, c_{i,t-1})$. The increment above the initial value, which we will denote $f_{i,t}(c_{i,t}, c_{j,t-1})$, is a continuous function of $c_{i,t}$ and $c_{j,t-1}$ only. From T-i(b), the increment does not depend on $\overline{p_{j,t}}$.

Feature (T-iv(b)): i's pricing function is continuous and strictly increasing in $\overline{p_{j,t}}$, and feature (T-iv(c)): i's pricing function is continuous and strictly increasing in (i's perception of) $c_{j,t-1}$.

Proof: The equilibrium price functions have the form

$$p^*_{i,t} = g_{i,t}(c_{i,t-1}, c_{j,t-1}) + f_{i,t}(c_{i,t}, c_{j,t-1})$$

where, as already shown, $g_{i,t}(c_{i,t-1}, c_{j,t-1})$ is the solution to

$$g_{i,t}(c_{i,t-1}, c_{j,t-1}) = \frac{a_i}{2b_{1,i}} + \frac{c_i}{2} + \frac{b_{2,i}}{2b_{1,i}} \overline{p_{j,t}}(c_{i,t-1}, c_{j,t-1})$$

which is increasing and continuous in $\overline{p_{j,t}}$. From the perspective of firm i, $\overline{p_{j,t}}$ is equal to

$$\overline{p_{j,t}} = \frac{a_j}{2b_{1,j}} + \frac{c_j}{2} + \frac{b_{2,j}}{2b_{1,j}} \overline{p_{i,t}} + \int_{\underline{c_j}}^{\overline{c_j}} f_{j,t}(c_{j,t}, c_{i,t-1}) \Psi_j(c_{j,t}|c_{j,t-1}) dc_{j,t}$$

where the continuity of the increment f and the conditional density $\Psi_j(c_{j,t}|c_{j,t-1})$, and the properties that (i) $f_{j,t}(c_{j,t}, c_{i,t-1})$ is increasing in $c_{j,t}$, and (ii) the integral is increasing in $c_{j,t-1}$ means that $\overline{p_{j,t}}$ is continuous and increasing in $c_{j,t-1}$, holding $\overline{p_{i,t}}$ fixed. But as $\overline{p_{i,t}}$ is also increasing, and continuous, in $\overline{p_{j,t}}$ and vice-versa, both pricing functions will also be increasing and continuous in both $c_{i,t-1}$ and $c_{j,t-1}$.

Strategies in Period T.

It remains to show that strategies in the final period have a form that will lead to the type of separating equilibrium strategies described above in period $T-1$. The required features are that:

- the period T equilibrium pricing function of firm i is continuous in $c_{i,T}$, $c_{i,T-1}$ and $c_{j,T-1}$; and,
- the expected value $\overline{p_{i,T}}$ is increasing in $c_{j,T-1}$.

In period T, both firms will use static optimal strategies given their beliefs about their rival's previous price. Therefore

$$p^*_{i,T} = \frac{a_i}{2b_{1,i}} + \frac{c_{i,T}}{2} + \frac{b_{2,i}}{2b_{1,i}} \overline{p_{j,T}}$$

where

$$\overline{p_{j,T}} = \frac{a_j}{2b_{1,j}} + \frac{b_{2,j}}{2b_{1,j}} \overline{p_{i,t}} + \frac{E(c_{j,T}|c_{j,T-1})}{2}$$

and solving these equations simultaneously gives

$$\overline{p_{j,T}} = \frac{\left(\frac{a_j}{2b_{1,j}} + \frac{a_i b_{2,i}}{4b_{1,j}b_{1,i}}\right) + \frac{b_{2,j}E(c_{i,T}|c_{i,T-1})}{4b_{1,j}} + \frac{E(c_{j,T}|c_{j,T-1})}{2}}{\left(1 - \frac{b_{2,i}b_{2,j}}{4b_{1,i}b_{1,j}}\right)}$$

so

$$p^*_{i,T} = \frac{a_i}{2b_{1,i}} + \frac{c_{i,T}}{2} + \frac{b_{2,i}}{2b_{1,i}} \left(\frac{\left(\frac{a_j}{2b_{1,j}} + \frac{a_i b_{2,i}}{4b_{1,j}b_{1,i}}\right) + \frac{b_{2,j}E(c_{i,T}|c_{i,T-1})}{4b_{1,j}} + \frac{E(c_{j,T}|c_{j,T-1})}{2}}{\left(1 - \frac{b_{2,i}b_{2,j}}{4b_{1,i}b_{1,j}}\right)} \right).$$

Given the form of Ψ_i and Ψ_j (A4), $p^*_{i,T}$ will be continuous in $c_{i,T}$, $c_{i,T-1}$ and $c_{j,T-1}$, and $\overline{p_{j,T}}$ is increasing in $c_{i,T-1}$, as required.

D Empirical Application: The Effects of the MillerCoors Joint Venture

This Appendix describes the data used in our analysis, as well as additional analysis that is not presented in the text. Readers are referred to Miller and Weinberg (2017) for more background on the JV as well as more details concerning the data/sample selection etc..

D.1 The Joint Venture.

The MC JV, announced in October 2007, effectively merged the U.S. brewing, marketing and sales operations of SABMiller (Miller) and MolsonCoors (Coors), the second and third largest U.S. brewers. The Department of Justice (DOJ) decided not to challenge the transaction in June 2008 because it expected "large reductions in variable costs of the type that are likely to have a beneficial effect on prices".[54] For example, the JV was expected to lower transportation costs by producing Coors products at Miller breweries around the country. Ashenfelter, Hosken, and Weinberg (2015) provide evidence that transportation efficiencies were realized.

MW show that, at a national level, the real prices (i.e., deflated by the CPI-U price index) of the most popular domestic brands, such as Bud Light (BL), Miller Lite (ML) and Coors Light (CL), increased after the JV, relative to the prices of imported brands, such as Corona Extra and Heineken, which MW use as controls for industry-wide cost shocks. Regressions in Appendix D.4 quantify these price increases to lie between 40 cents and a dollar per 12-pack, or 3%-6%, depending on the specification. We will proceed assuming that MW's interpretation that the relative price increase was a causal anticompetitive effect of the JV is correct.[55]

An important feature of the relative price change is that AB's prices increased as much as those of Miller and Coors. If AB's marginal costs were unaffected by the JV, this pattern is inconsistent with static CI Nash pricing, as a static best response function would predict that AB should have responded to any JV price increase by raising its prices by a smaller amount.

[54]Department of Justice press release, 5 June 2008.

[55]This interpretation is complicated by how the Great Recession may have affected demand and the fall in the deflator, from 220.0 in July 2008 to 210.2 in December 2008, at exactly the same time that the merger was being consummated.

D.2 IRI Data.

The data comes from the beer category of the IRI Academic Dataset (Bronnenberg, Kruger, and Mela (2008)). The underlying data is at the weekly UPC-store-level from 2001 to 2011. We only use data from grocery stores.

We use different samples at different points of our analysis. For example, when we extend MW's demand and conduct parameter analysis, we use their samples, whereas when we calibrate our model we use a sample that we view as appropriate. For example, MW exclude sales of cans and bottles in 18-packs. These are unimportant for most brands, but account for more than 20% of volume sold for the three domestic flagship brands (Bud Light (BL), Miller Lite (ML) and Coors Light (CL)) so we do not want to exclude them. We also stop our pre-JV sample at the time that the JV was announced, excluding the period of the DOJ's investigation when ML prices dropped dramatically.

D.2.1 Data Selection for the Demand and Conduct Analysis (presented in Appendices D.8 and D.9).

We follow MW in using the following selection of data.

- **selection of markets:** 39 geographic (IRI defined) regional markets excluding (e.g., because they lack other types of data that will be used in demand estimation, or are viewed as having too few beer sales) the following markets with some stores selling beer in the data: Harrisburg/Scranton; Philadelphia; Providence RI; Tulsa; Minneapolis-St. Paul; Oklahoma City; Salt Lake City; Kansas City; New England; Pittsfield; Eau Claire, WI.

- **brands:** 13 brands, which are BL, ML, CL, Budweiser, Miller Genuine Draft, Miller High Life, Coors, Corona Extra, Corona Light, Heineken, Heineken Premium Light, Michelob Ultra, Michelob Light.

- **pack sizes:** packages of cans and glass bottles containing the equivalent of 6, 12, 24 and 30 12oz. servings. 24 and 30-packs are aggregated into a single "large" size. Prices are calculated as total dollars sold divided by volume in 12-pack equivalents.

- **product:** a product is a brand × pack size (6-pack, 12-pack, "large") combination.

- **time periods:** for demand and supply estimation, data from January 2005 to December 2011 is used, but months from June 2008 to May 2009, i.e., the period immediately after the JV was consummated, are excluded. Monthly data is created by allocating individual days within a week to their correct month, and assuming that sales within a week are spread equally across the days in the week, before aggregating to the monthly level.

- **distances and diesel prices:** we use MW's estimated distance from the brewery or port (for Heineken) to the market, measured in thousands of miles. Monthly diesel prices come from the U.S. Energy Information Administration.

- **income data:** the random coefficients models are estimated using data on household income taken from the 2005-2011 PUMS samples of the American Community Survey (ACS). We use the same samples as MW to estimate demand.

- **deflator:** when using real prices, or real diesel prices, they are deflated to January 2010 levels using the CPI-U All Urban Consumers-All Items price index.

The following additional variables are defined:

- **market size:** for each market, market size is defined as 150% of the maximum of the total sales, measured in 12-pack equivalents, of all of the brands listed above plus 23 others (including popular brands such as Busch and Busch Light) in the package sizes/types that are being used. When we estimate demand using weekly data, we use an alternative definition that defines demand as 150% of the sum of the maximum sales across the stores observed in the sample that week.

- **distance measure:** the distance measure is constructed by multiplying deflated diesel prices by the driving distance from the brewery, or port in the case of Heineken, to the market.

- **demand instruments:** to estimate demand it is necessary to define instruments for a product's price and its share of volume sold amongst the products in its nest. MW use the following instruments:

  - the product's own distance measure (iv-1)

- the sum of the distance measures for all of the products in the nest (iv-2)
- the number of products in the nest (iv-3)
- a dummy for domestic products after the JV (iv-4)
- (iv-2) and (iv-3) interacted with a dummy for products produced by Miller, Coors, AB or MillerCoors
- (iv-2) and (iv-3) interacted with a dummy for products produced by AB

When we estimate demand allowing for a flagship nest and an "other brand" nest, (iv-1) and (iv-4) are interacted with a dummy for flagship products, and the other instruments are defined at the nest level (e.g., adding over all products in the same nest, rather than all products). However, all three package sizes are available for all flagship products in all markets, so, for the flagship nest, the (iv-3) instruments are dropped due to collinearity.

D.2.2 Data Selection for the Calibration of Our Model.

For our calibration we depart from MW's selection in the following aspects.

- **selection of markets:** we use observations from all market-weeks where we observe the flagship brands being sold in at least 5 stores. This gives us 45 markets before the JV, although some markets do not meet the criteria in some weeks. The markets that are added back are: Eau Claire, Kansas City, Minneapolis, New England, Oklahoma City, Salt Lake City. Boston never meets the 5 store criterion after the JV so it is excluded from our estimates of post-JV price dynamics.

- **pack sizes:** packages of cans and glass bottles containing the equivalent of 6, 12, 18, 24 and 30 12oz. servings. These sizes are treated separately, but prices are converted into 12-pack equivalents.

- **time periods:** we use the months from January 2001 to October 2007 for the pre-JV period. The months after May 2009, until December 2011, are the post-JV period.

Table D.1: Highest-Selling Beer Brands in 2007 with Ownership, Share and Average Nominal Retail Prices per 12-Pack.

			2007			2011	
Brand	Company	Packs	% 18+	Mkt. Share	Price	Mkt. Share	Price
Bud Light*,†	AB	10	72.5%	15.7%	$8.29	15.7%	$8.92
Miller Lite*,†	M	10	75.1%	10.0%	$8.11	8.4%	$8.73
Coors Light*,†	C	10	74.8%	8.3%	$8.36	9.4%	$8.98
Budweiser†	AB	10	70.8%	7.7%	$8.30	6.5%	$9.00
Corona Extra$^{\dagger,\diamond}$	GM	5	15.6%	4.1%	$13.88	3.9%	$13.46
Natural Light*	AB	7	68.6%	3.9%	$6.01	3.2%	$7.15
Busch Light*	AB	9	78.4%	2.8%	$6.07	2.5%	$6.96
Miller High Life†	M	9	54.1%	2.4%	$6.33	2.2%	$7.21
Heineken$^{\dagger,\diamond}$	H	7	12.8%	2.3%	$14.06	2.3%	$13.86
Miller Genuine Draft†	M	10	67.0%	2.3%	$8.26	1.3%	$8.94
Michelob Ultra*,†	AB	9	27.4%	2.1%	$10.05	2.4%	$10.51
Busch	AB	9	70.0%	1.9%	$6.08	1.6%	$7.05
Keystone Light*	C	6	81.4%	1.4%	$5.83	1.5%	$7.03
Budweiser Select	AB	9	62.0%	1.3%	$8.37	0.7%	$8.76
Milwaukee's Best Light*	M	6	66.8%	1.3%	$5.37	0.8%	$6.19
Corona Light$^{*,\dagger,\diamond}$	GM	3	2.3%	1.2%	$14.23	1.3%	$13.79
Tecate$^{\diamond}$	H	7	66.3%	1.2%	$8.65	1.2%	$9.04
Natural Ice	AB	7	51.3%	1.1%	$5.96	0.9%	$7.19
Pabst Blue Ribbon	SP	9	49.3%	1.0%	$6.26	1.4%	$7.53
Milwaukee's Best	M	5	61.8%	0.8%	$5.46	0.4%	$6.46
Coors†	C	10	73.3%	0.8%	$8.44	1.0%	$8.84
Michelob Light*,†	AB	7	29.3%	0.7%	$9.76	0.3%	$10.72
Heineken Prem. Light$^{*,\dagger,\diamond}$	H	5	1.9%	0.6%	$14.28	0.5%	$14.18

Notes: the table lists the 20 highest-selling brands plus additional brands in MW's sample. Market shares and prices are based on all units sold in packs equivalent to 6, 12, 18, 24, 30 and 36 12oz servings. "Packs" is the number of 2007 bottle/can-pack size combinations for 6, 12, 18, 24 and 30 packs, as 36 packs are rare. "% 18+ " is the percentage of 2007 volume sold in the packs of more than 18 cans or bottles. 2007 companies are: AB=Anheuser-Busch, M=SABMiller, C=MolsonCoors, GM=Grupo-Modelo, H=Heineken, SP=S&P. Prices are nominal prices per 12-pack equivalent (i.e., total dollars sold in all pack sizes divided by total volume in 144oz. units). *=light beers, †=included in MW's sample, $^{\diamond}$=imports.

D.3 Brand Shares and Retail Prices.

Table D.1 lists the 20 brands with the largest sales by volume in 2007, together with additional brands that MW include in their analysis. The table lists market shares and average nominal prices (per 144 oz, the volume in a standard 12-pack) in 2007 and 2011.

Most domestic brands are differentiated from imports by being sold primarily in larger packs and at lower prices. The relative prices of domestic brands increased after 2007, but, although CL gained some share at ML's expense, the domestic brewers' market shares remained stable. For example, AB's volume share was 41.3% in 2007, 41.5% in 2009 and 39.6% in 2011, with light beer shares of 50.0%, 50.8% and 50.6% respectively.[56]

D.4 Effects of the Joint Venture on Prices.

MW present estimates of the effects of the joint venture on prices. We present complementary estimates here, which can be compared to the price increases predicted by our calibrated model.

An observation in our analysis is a brand-market-month, where real prices are calculated at the brand level by adding up the total sales in package sizes equivalent to packs of 6, 12, 18, 24, 30 or 36 12oz. containers (we include 36-packs in this regression where they are available, although they account for a small proportion of sales). The sample contains the following brands: BL, ML and CL (i.e., the domestic flagship brands), Corona Extra and Heineken which we will treat as providing controls for industry-wide shocks, as MW assume. The sample runs from 2001 to 2011, and includes the period immediately before and following the JV. We consider prices defined using all store-UPC-week observations in the appropriate sizes, and prices that are defined excluding store-UPC-week observations that are identified as being sold at temporary price reduction prices. We use both definitions as our analysis of price dynamics will use price series where price reductions are removed.

Table D.2 presents the results from six specifications that differ depending on whether price reductions are included, we use prices in levels or logs and whether brand-time trends are included. The reported coefficients are the coefficients on Post-JV dummies for the domestic flagship brands, so that they measure the increase in real prices relative to the two imported

[56]Appendix D.5 presents a figure showing the evolution of market shares over this period. The post-JV decline in the shares of several non-flagship domestic brands reflected a continuation of pre-existing trends.

Table D.2: Estimates of the Effects of the Joint Venture on Prices.

	(1)	(2)	(3)	(4)	(5)	(6)
	$ Price/ 12 Pack	Log(Price/ 12 Pack)	$ Price/ 12 Pack	Log(Price/ 12 Pack)	$ Price/ 12 Pack	Log(Price/ 12 Pack)
Price Reductions	incl.	incl.	incl.	incl.	excl.	excl.
Post-JV Brand Dummies						
Bud Light	0.853	0.046	0.428	0.046	0.485	0.032
	(0.049)	(0.005)	(0.064)	(0.005)	(0.080)	(0.007)
Miller Lite	1.024	0.065	0.415	0.045	0.492	0.034
	(0.058)	(0.006)	(0.071)	(0.006)	(0.070)	(0.006)
Coors Light	0.945	0.056	0.438	0.048	0.542	0.040
	(0.060)	(0.006)	(0.068)	(0.006)	(0.076)	(0.007)
Brand Time Trends	N	N	Y	Y	Y	Y
Observations	25,740	25,740	25,740	25,740	25,740	25,740
R^2	0.971	0.973	0.972	0.973	0.970	0.970

Notes: the reported coefficients are on domestic brand × post-JV interactions. The brands included are those listed, plus Corona Extra and Heineken. Observations at the brand-market-month level, aggregating across packages containing the equivalent of 6, 12, 18, 24, 30 or 36 12oz. containers in cans or glass bottles. All specifications include market-brand and time period fixed effects. Standard errors in parentheses clustered on the market.

brands. The estimated price increases vary across the columns, but lie in the range from just over 40 cents to one dollar, or 3% to 6%, and the price increases are smaller when we include brand-specific time trends.

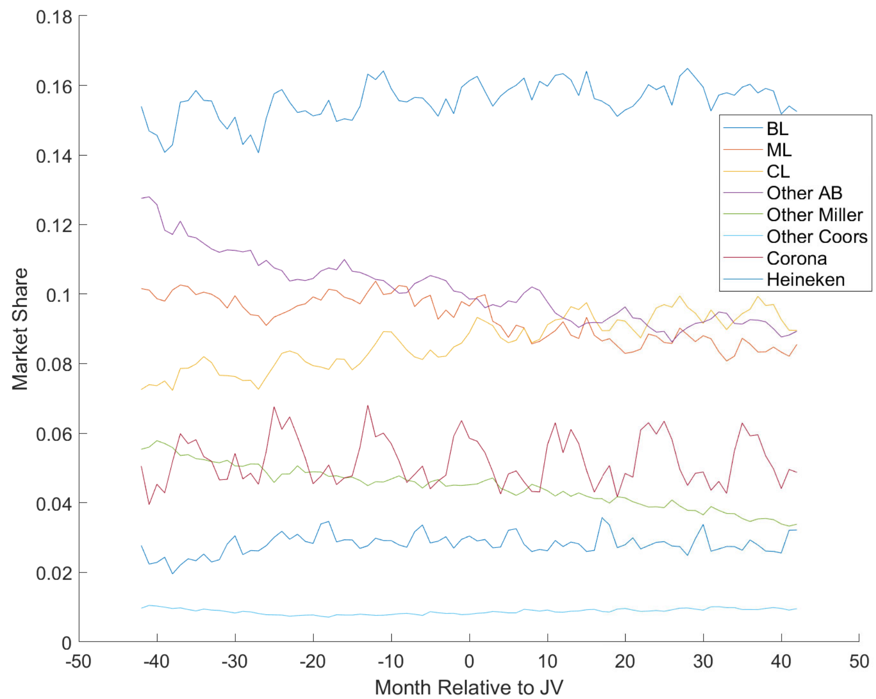
We will assume that these relative price changes reflect causal, anticompetitive effects of the JV, which can be compared to the predictions of the effects of an unanticipated, exogenous merger in our model. Of course, this interpretation does depend on assumptions, in particular the validity of using the prices of imported brands as controls for cost changes.

D.5 Changes in Market Shares Around the Joint Venture.

Our preferred demand system for the calibration assumes that there is limited substitution between the flagship domestic brands, brands that are not owned by the leading domestic firms or the outside good of not drinking beer, so that post-JV price increases of the observed size do not reduce demand of the leading domestic brands very much. This is consistent with some of our estimates in Table D.4, although MW's specifications imply more substitution.

Figure D.1 shows the volume-based market shares of the different brands included in the

Figure D.1: Brand Market Shares Around the Joint Venture



Notes: Budweiser, Michelob Ultra and Michelob Light aggregated into "Other AB"; Miller Genuine Draft and Miller High Life aggregated to "Other Miller"; Coors is "Other Coors"; Heineken and Heineken Premium Light are "Heineken" and Corona Extra and Corona Light are "Corona". Shares based on volume sold in packages equivalent to 6, 12, 18, 24, 30 and 36 12oz containers.

demand analysis (for this purpose, we define market share based on the shares of **all** beers in the IRI data, not just the ones that MW include in their demand model). We aggregate non-flagship brands based on their pre-JV ownership. The main feature of the figure is that while the real prices of the flagship brands and the other domestic brands increase after the JV, brand market shares are stable, except for CL gaining share at the expense of ML, a change that does not appear to be driven by prices. The shares of non-flagship Miller and AB brands do fall after the JV, but this appears to reflect trends that existed before the JV. Imported beers do not appear to gain share.

D.6 Price Correlations Across Brands Before and After the JV.

Our model assumes that each firm sets a single price per period, including MC after the JV, whereas the domestic brewers have large portfolios of brands that are sold in many different packages. In this Appendix we show that the prices of brands sold by the same brewer are highly correlated, and that Miller and Coors brand prices are more correlated after the JV. This provides some comfort that our simplifying assumption is not too misleading.

Table D.3: Cross-Brand Correlations in Prices for 12-Packs

		Pre-JV					
		(1)	(2)	(3)	(4)	(5)	(6)
(1)	Bud Light	1					
(2)	Miller Lite	0.891	1				
(3)	Coors Light	0.891	0.889	1			
(4)	Budweiser	0.994	0.892	0.893	1		
(5)	Miller Genuine Draft	0.872	0.973	0.870	0.872	1	
(6)	Coors	0.804	0.812	0.916	0.807	0.804	1
		Post-JV					
		(1)	(2)	(3)	(4)	(5)	(6)
(1)	Bud Light	1					
(2)	Miller Lite	0.857	1				
(3)	Coors Light	0.874	0.967	1			
(4)	Budweiser	0.995	0.856	0.872	1		
(5)	Miller Genuine Draft	0.840	0.957	0.940	0.839	1	
(6)	Coors	0.825	0.934	0.959	0.824	0.916	1

Notes: the correlations are for brand-market-week average prices of 12-packs, before the announcement of the JV and after its consummation. Average prices are calculated including price reductions. Correlations for brands with the same owner are slightly higher if price reductions are excluded.

Table D.3 reports the correlations of market-week prices of 12-packs of the flagship brands, plus Budweiser, Miller Genuine Draft and Coors, before and after the JV. It is noticeable that the prices of products with the same owner (e.g., BL and Budweiser) are highly correlated and that the prices of Miller and Coors products become more correlated after the JV.

The reported correlations are high partly because beers retail at different prices in different markets. We can also calculate correlations by regressing the price of one brand on the price of another brand, and market and week fixed effects. These results also show significant increases in correlations of Miller and Coors products after the JV: for example, the coefficient on the

CL price when the ML price is the dependent variable increases from 0.68 before the JV to 0.84 after the JV. Patterns in the table and the regressions are similar if we use prices defined to exclude temporary price reductions.

D.7 Price Series of Domestic Flagship 12-Packs Nationally, and in Los Angeles and Seattle.

Figure 6 in the main text plots the monthly time series of average nominal retail prices for 12-packs of Bud Light, Miller Lite and Coors Light. The averages are calculated by dividing total dollar sales (excluding sales at temporary store price reductions) by the number of units sold. Price volatility is a clear feature of the time-series, and it is arguably a clearer feature than the post-JV price increase. This motivates our use of a model where price volatility is a feature of equilibrium pricing.

However, one might be concerned that volatility is partly driven by how average prices have been calculated. Therefore, Figures D.2-D.4 show price series calculated in three different ways:

- nominal prices, including temporary store price reductions;
- real (deflated) prices, excluding temporary store price reductions; and,
- nominal average prices where the same weight (rather than volume share weights) are placed on each store that is in the sample.

The third approach is motivated by the possibility that, at the store-level, prices do not change, but that volatility in share-weighted average prices is due to the number of units sold in high-priced and low-priced retail stores varying over time. While the different calculations do change the level of average prices, volatility remains.

Figure D.2: Real Prices for 12-Packs of the Domestic Flagship Brands Nationally and in Two Regional Markets Around the JV. Prices are deflated monthly average prices excluding sales indicated to be made at temporary price reductions.

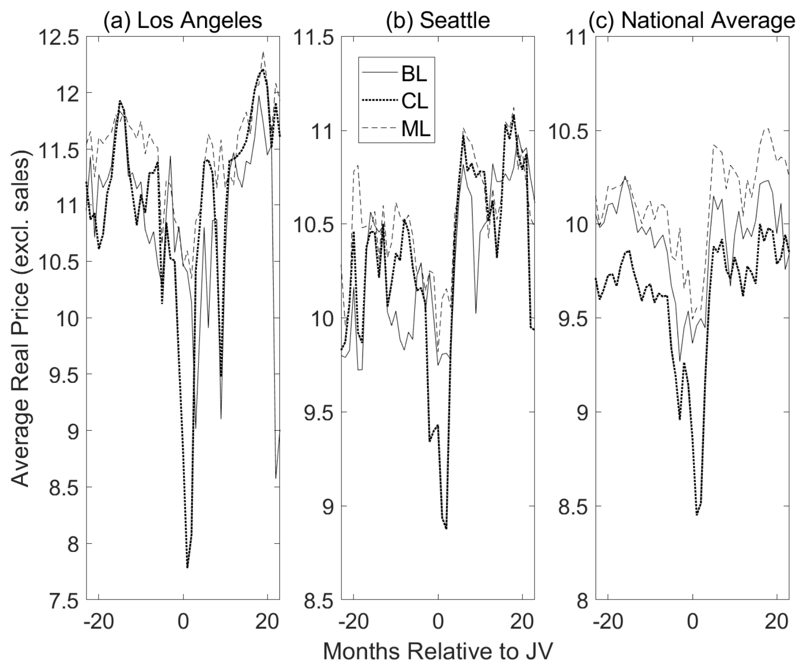


Figure D.3: Nominal Prices for 12-Packs of the Domestic Flagship Brands Nationally and in Two Regional Markets Around the JV. Prices are monthly average prices including sales indicated to be made at temporary price reductions.

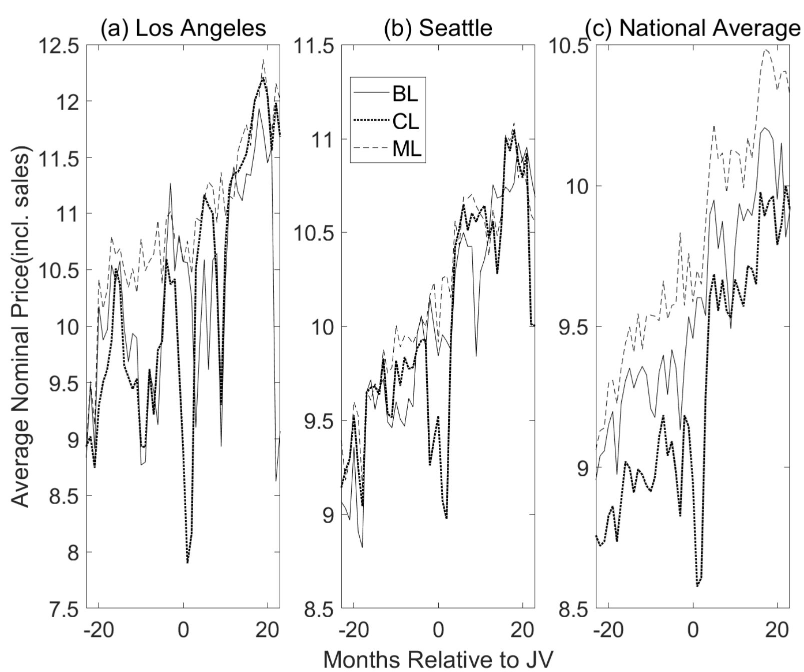
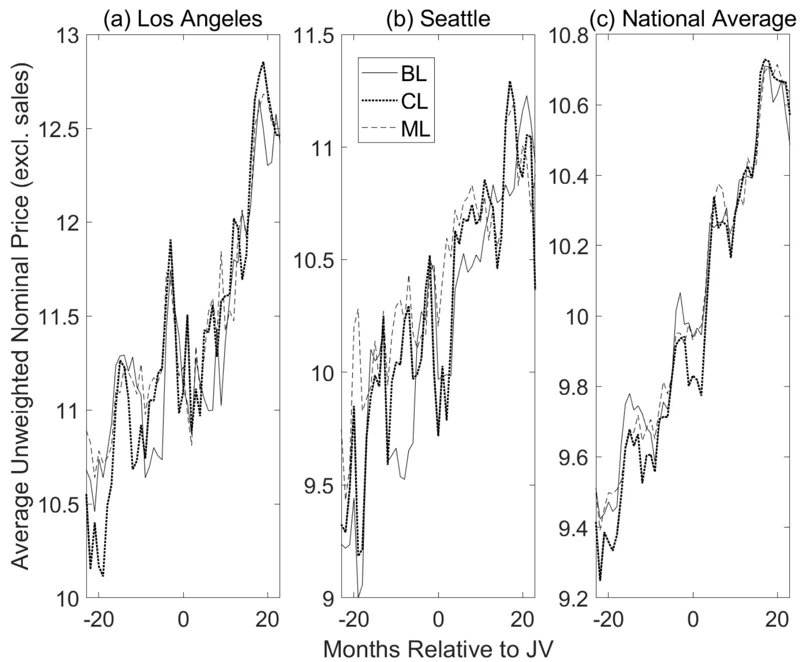


Figure D.4: Nominal Average Store Prices for 12-Packs of the Domestic Flagship Brands Nationally and in Two Regional Markets Around the JV. Reported values are unweighted average monthly prices average across stores, where monthly average prices are calculated excluding sales made at temporary price reductions.



D.8 Demand.

We estimate several demand specifications using the same selection of data that MW use. We use the demand parameters as an input into our re-examination of the conduct and supermarkup models presented by MW and MSW, and to support the simpler parameterization of demand that we use when calibrating the supply-side parameters of our model.

Table D.4 reports five sets of demand estimates (the first three will be used in Appendix D.9). For these specifications, we follow MW as closely as possible in the choice of data, instruments and controls, except that we use optimal GMM for the nested logit models as doing so affects the estimates.[57] The first three columns contain one nested logit specification, using monthly data, and two random coefficients nested logit (RCNL) specifications, where the 13 MW brands are all included in a single inside nest, and preferences vary with income. The remaining columns estimate nested logit models using monthly and weekly data (we will use weekly price changes when estimating the cost parameters) where the flagship BL, ML

[57]None of the specifications yield exactly the same estimates as MW, although the monthly RCNL coefficients are almost identical.

and CL products are grouped into a “flagship nest”, and the remaining products are placed in an “other beer” nest with a different nesting coefficient. The flagship nesting coefficients are larger, consistent with these brands being particularly close substitutes.

The table reports several implied statistics for each specification, including the average (across market-time periods) ML brand elasticity (i.e., the effect on demand when all ML prices increase), the proportion of lost demand that switches to other flagship products when a flagship price is increased, and the average, across pre-JV observations, predicted change in flagship sales when the prices of all domestic products increase by 75 cents, which is within the range of the observed post-JV price change.

The statistics vary across the specifications. Recall that domestic/flagship brand market shares fall very little, if at all, after the JV in spite of the price increase (Appendices D.3-D.5). Among the five specifications, this is most consistent with the estimates in columns (4) and (5). In our calibration, we will therefore assume a price elasticity and a diversion rate which is consistent with the estimates in these columns.

D.9 Tests of Collusive Models of Pricing in the Beer Industry.

Some people have suggested that, even if our model can explain why prices rose after the MC JV, CI theories of tacit collusion provide pre-existing and satisfactory explanations. This is true in the sense that folk theorems imply that one could likely construct some CI tacit collusion model that would be able to explain any feature of prices almost perfectly. However, the specific models advanced by MW and MSW can be tested. MW provide a supply-side framework where they account for the increase in prices by a “change in conduct”, from assumed Bertrand Nash pricing before the JV to partial joint-profit maximization after the JV, where the latter type of conduct is viewed as reflecting tacit collusion. We will show that, using alternative identifying assumptions, we can typically reject the hypothesis of Nash pricing before the JV while, contrary to MW’s assumption, not necessarily rejecting a hypothesis that there was no change in conduct after the JV. MSW propose a specific model of tacit collusion where the domestic firms charge market-year specific markups above Bertrand Nash prices (supermarkups). We show that the data rejects this model.

Of course, these findings do not indicate that a model based on asymmetric information

Table D.4: Estimates of Demand

	(1)	(2)	(3)	(4)	(5)
	Nested Logit	RCNL	RCNL	Nested Logit	Nested Logit
Nests	All Beer	All Beer	All Beer	Flagship/Other	Flagship/Other
Data Freq.	Monthly	Monthly	Quarterly	Monthly	Weekly
Real Price Coefficient	-0.056	-0.083	-0.099	-0.073	-0.047
(2010 dollars)	(0.017)	(0.014)	(0.014)	(0.018)	(0.011)
Nesting Coefficients					
Single All Brand Nest	0.741	0.838	0.831	-	-
	(0.051)	(0.039)	(0.039)		
Two Nests					
Domestic Flagship	-	-	-	0.838	0.898
				(0.049)	(0.040)
Other Brands	-	-	-	0.634	0.815
				(0.047)	(0.037)
Income Coefficients (RCNL Models)					
*constant	-	0.014	0.014	-	-
		(0.005)	(0.005)		
*price	-	0.001	0.001	-	-
		(0.000)	(0.000)		
*calories	-	0.004	0.004	-	-
		(0.002)	(0.002)		
Median Product Elasticity	-2.31	-4.71	-5.41	-2.51	-3.12
Mean ML Brand Price Elasticity	-1.66	-3.68	-4.22	-3.06	-3.09
Mean Flagship Diversion	0.41	0.48	0.47	0.83	0.90
% Change in Flagship Sales Given 75¢ Domestic Price Rise	-5.20%	-8.24%	-9.65%	-4.30%	-2.20%
Observations	94,656	94,656	31,777	94,656	405,004

Notes: all specifications include time period and product (brand*size) fixed effects, and use data from Jan 2005 to Dec 2011, excluding June 2008 to May 2009. All estimates use two-step optimal GMM. Instruments are the same as in MW for the relevant specification, apart from the two nest models where we define instruments for the number and distance measures for other products based on products in the same nest, and interact instruments with a flagship brand dummy. Market size is defined as 50% more than the highest sales observed in the geographic market for monthly and quarterly specifications. For the weekly specifications it is estimated as 50% more than the sum of the highest sales from stores observed in the scanner data that week. ML Brand Elasticity reflects the change in ML sales when the prices of all ML products are increased. Mean Flagship Diversion is the average proportion of lost sales that go to other flagship products (i.e., BL, ML and CL products) when the price of a flagship product is increased. The change in flagship sales after a 75 cent price rise is the average across pre-JV observations change in total flagship sales when the prices of all domestic products are increased by 75 cents. Standard errors, clustered on the geographic market, in parentheses.

and signaling is the "correct" model, because CI is assumed when deriving the first-order conditions that are estimated. However, the estimates do imply that marginal costs are serially correlated and quite volatile, a feature that plays an important role in our model. We have also experimented with estimating conduct-model equations using data generated from the asymmetric information model examples presented in Section 3. We find conduct parameter estimates that are broadly consistent with those that we estimate during observed market data in this Appendix, i.e., estimated conduct parameters are significantly greater than zero, and may fall slightly, increase or stay roughly the same after a simulated merger.[58]

D.9.1 The Conduct Parameter Framework.

Our tests extend MW's conduct parameter framework. The framework assumes that pricing is characterized by stacked static, CI first-order conditions

$$\left(\Omega_{mt} \circ \left[\frac{\partial q_{mt}(p_{mt}, \theta^D)}{\partial p_{mt}}\right]\right)(p_{mt} - c_{mt}) + q_{mt}(p_{mt}, \theta^D) = 0,$$

where p_{mt}, q_{mt} and c_{mt} are vectors of prices, quantities and (constant) marginal costs and $\frac{\partial q_{mt}(p_{mt},\theta^D)}{\partial p_{mt}}$ is a matrix of demand derivatives.

Ω_{mt} is the "conduct" matrix, with (row i, column j) element $\Omega_{i,j}$. $\Omega_{i,j} = 1$ if products i and j are owned by the same firm. Under static Nash pricing, all other elements of Ω_{mt} are zero. MW's baseline specification assumes static Nash pricing before the JV, but allows $\Omega_{i,j} = \kappa$ after the JV if i and j are owned by different domestic brewers. $\kappa = 1$ is consistent with joint profit-maximization, while $0 < \kappa < 1$ could be interpreted as reflecting partial internalization of pricing externalities.

Given demand estimates, MW estimate the post-JV κ using equations

$$p_{mt} = W_{mt}\gamma - \left(\Omega_{mt}(\kappa) \circ \left[\frac{\partial s_{mt}(p_{mt}, \theta^D)}{\partial p_{mt}}\right]\right)^{-1} s_{mt}(p_{mt}) + \nu_{mt}. \tag{5}$$

where $c_{imt} = W_{imt}\gamma + \nu_{imt}$ and W includes time, product (brand-size) and geographic market

[58] One might ask why we do not do this exercise with data simulated from our calibrated model. The answer is that identification of separate pre- and post-merger conduct parameters relies on cross-market or across-time variation in the degree of substitution between different brands. However, in order to have a manageable computational burden, our calibrated model assumes a single average or representative market.

fixed effects; a "distance measure" that multiplies distance to the brewery or port with real diesel prices, and allows for the JV to realize transportation cost efficiencies by reducing distances from the market for MC products; and, a dummy for MC products after the JV to allow for an additional efficiency. The identifying assumption/exclusion restriction is that the JV is assumed not to affect AB's marginal costs. The instruments are the variables in W and a dummy for domestic products after the JV. The post-JV κ is identified by how much more AB's prices increase than the increase that can be rationalized as the static best response implied by the static CI first-order conditions.

MW's single exclusion restriction implies that they cannot estimate separate pre- and post-JV κs or test whether a change in conduct is the source of the price increase.[59] Our approach is to estimate separate pre- and post-JV κs by adding additional instruments and controls. We continue to assume that imported brands use Nash pricing and that $\Omega_{i,j} = 1$ when i and j have the same owner. Note, however, that we will only use the model to test MW and MSW's assumptions and we will not interpret positive κs as evidence of "collusion". As shown by Corts (1999), some forms of tacit collusion may be consistent with estimates of κ that are less than or equal to zero, and, as we noted above, one usually estimates positive and statistically significant κs using data simulated from our model even though there is no collusion.

Our specifications include separate pre- and post-JV product and market fixed effects in W, so changes in price levels after the JV do not identify conduct. To understand our choice of instruments, consider the first-order condition for product i owned by AB

$$p_{imt} = W_{imt}\gamma + \frac{q_{imt}}{\frac{\partial q_{imt}}{\partial p_{imt}}} + \sum_{\substack{j \in AB \\ j \neq i}} \frac{\frac{\partial q_{jmt}}{\partial p_{imt}}}{\frac{\partial q_{imt}}{\partial p_{imt}}}(p_{jmt} - c_{jmt}) + \kappa \sum_{k \in M,C} \frac{\frac{\partial q_{kmt}}{\partial p_{imt}}}{\frac{\partial q_{imt}}{\partial p_{imt}}}(p_{kmt} - c_{kmt}) + \nu_{imt}.$$

Valid instruments will be correlated with $\sum_{k \in M,C} \frac{\frac{\partial q_{kmt}}{\partial p_{imt}}}{\frac{\partial q_{imt}}{\partial p_{imt}}}(p_{kmt} - c_{kmt})$ (i.e., the incremental effect of a change in i's price on a rival's profits), and uncorrelated with the cost unobservable ν_{imt}. We will use alternative instruments in the specifications below.

Table D.5: Testing Alternative Models Using a Generalized Conduct Parameter Framework

	Tests of MW Conduct Model						Tests of MW & MSW Models			Test of MSW
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Demand Model	NL	RCNL	RCNL	NL	RCNL	RCNL	NL	RCNL	RCNL	RCNL
(all one nest)	Monthly	Monthly	Quarterly	Monthly	Monthly	Quarterly	Monthly	Monthly	Quarterly	Quarterly
Domestic Firms										FY06, FY07
Pre-JV Conduct	0.274	0.322	0.198	0.340	0.263	0.238	0.958	0.909	0.913	0.977, 0.924
	(0.098)	(0.193)	(0.221)	(0.057)	(0.166)	(0.147)	(0.005)	(0.016)	(0.024)	(0.007),(0.013)
										FY10, FY11
Post-JV Conduct	0.723	0.651	0.573	0.688	0.767	0.717	0.951	0.914	0.921	0.976, 0.933
	(0.124)	(0.146)	(0.127)	(0.144)	(0.056)	(0.062)	(0.009)	(0.014)	(0.013)	(0.011),(0.015)
p-value diff.	0.004	0.013	0.017	0.000	0.000	0.001	0.483	0.638	0.558	-
Supermarkup	-						Domestic*Market*Fiscal Year Fixed Effects			
Controls							linear			non-linear
Excluded IVs	Domestic Rival Distance			Dom. Rival Distance, Dom. Rival ξs and Interactions			Dom. Rival Distance, Dom. Rival ξs and Interactions			
ML 12 Packs										
Pre-JV: Mean $\widehat{c_{imt}}$	\$2.37	\$5.93	\$6.75	\$1.99	\$6.13	\$6.65	-\$4.61	\$1.73	\$2.79	-\$1.41
Residual ρ	0.414	0.427	0.451	0.410	0.430	0.449	0.239	0.252	0.114	0.030
SD AR(1) res.	\$0.31	\$0.27	\$0.20	\$0.31	\$0.27	\$0.20	\$0.63	\$0.44	\$0.30	\$0.35
Post-JV: Mean $\widehat{c_{imt}}$	-\$1.34	\$4.20	\$5.40	-\$1.00	\$3.34	\$4.62	-\$4.39	\$1.81	\$2.93	-\$5.74
Residual ρ	0.431	0.488	0.403	0.433	0.485	0.417	0.224	0.436	0.050	0.006
SD AR(1) res.	\$0.43	\$0.33	\$0.26	\$0.41	\$0.37	\$0.28	\$0.56	\$0.42	\$0.29	\$0.43
Observations	94,656	94,656	31,777	94,656	94,656	31,777	94,656	94,656	31,777	31,777 total

Notes: Specifications estimated using 2-step GMM. The specifications in columns (1)-(9) contain time period fixed effects, and separate product and market fixed effects for before and after the JV, as well as the distance measure interacted with combinations of dummies for domestic products and periods after the JV. The specification in column (10) is estimated separately for each fiscal year (e.g., the FY06 year runs October 2005-September 2006), and the specification includes product, city and quarter fixed effects, the distance measure (interacted with a dummy for domestic products) as well as non-linear market fixed effects for the domestic products. Conduct parameters are reported for four fiscal years. The "residual ρ" statistics are the coefficients on lagged marginal costs (c_{imt-1}) from a regression of ML 12-pack marginal costs on their lagged values, market and time fixed effects. These regressions are estimated separately before and after the JV. The "SD AR(1) res." statistics are the standard deviation of the residuals from these regressions. Standard errors in parentheses clustered at the market level.

D.9.2 Results: Estimates of Conduct Before and After the JV.

The first six columns in Table D.5 report conduct coefficients for the columns (1)-(3) demand specifications in Table D.4.[60] Columns (1)-(3) use the distance measures of rivals as instruments, as they affect rivals' margins, and, as MW already assume that a product's own distance measure is uncorrelated with ν_{imt}, the additional assumptions required are minimal.[61] Columns (4)-(9) use additional instruments in the form of the average value of the demand unobservables (ξs) for rival brewers over either the pre- or post-JV period, and the interactions of these instruments with the distance instruments.[62] These additional instruments are valid if ν_{imt} is uncorrelated with the demand unobservables of rivals' products. This is a stronger assumption, although economists sometimes make an even stronger assumption that a product's own demand and marginal costs unobservables are uncorrelated in order to estimate demand (MacKay and Miller (2019)). Columns (7)-(9) include linear domestic-market-fiscal year fixed effects in W. These controls allow for possible correlations between local preferences and costs for domestic products as a group, and cause conduct to be identified only from within-market-year cross-brewer/-product variation. We will also use these specifications to test the MSW model, as we explain below.

We reject Nash pricing after the JV in all nine specifications. This is, of course, consistent with MW's interpretation that there was tacit collusion after the JV. However, all of the estimated pre-JV κs are positive, and some are significant, providing evidence against the assumption of static CI Nash pricing before the JV. The estimates in columns (1)-(6) are consistent with an increase in κ after the JV, but the estimates with market-year controls

[59] MW re-estimate the post-JV κ assuming, but not estimating, different pre-JV $\kappa \leq 0.5$. These estimates imply that κ rose after the JV, although by smaller amounts as the assumed pre-JV κ rises, as a pre-JV κ also implies that AB would increase its prices when MC benefits from an efficiency.

[60] We have also estimated specifications using the two (flagship/non-flagship) nest nested logit models, and specifications that estimate κs based only on the pricing of the flagship brands. These estimates lead us to reject Nash pricing behavior before the JV, and the pre- and post-JV parameters are closer than those in columns (1)-(6).

[61] There are eight excluded distance instruments. For AB products in market m and time t before the JV, the (m, t) distance measure for Miller and the (m, t) distance measure for Coors are instruments. For pre-JV Miller products, the distance measures for AB and Coors are instruments. For pre-JV Coors products, the distance measures for Miller and AB are instruments. For AB (MC) products in market m and time t after the JV, the (m, t) distance measure for MC (AB) is the instrument.

[62] Specifically, we calculate the average value of the demand residuals for the products sold by brewer b in market m either before or after the JV, and then construct eight instruments in the same way that we construct the instruments for distance. We average across periods because the demand unobservables are more variable than the distance measures.

suggest that conduct did not change, even though these estimates of κ are the most precise.

The plausibility of these static CI pricing models can also be assessed by looking at what they imply for marginal costs and synergies. The lower panel of Table D.5 reports average implied marginal costs for ML 12-packs. Less elastic demand and higher κ imply lower marginal costs, and the (1), (4) and (7)-(9) costs are implausibly/impossibly low. The remaining columns imply synergies for ML, which was being shipped the same distances before and after the JV in most markets, that are higher than the 17.5% synergy for ML and CL that we assumed for the column (1) specification of our model. Controlling for market and time effects, the implied ν_{imt}s (marginal cost residuals) are also serially correlated and quite volatile.[63] While cost volatility is certainly not inconsistent with CI, we view volatility as suggesting that an interpretation of the data as reflecting tacit collusion requires a very strong CI assumption: if CI is not satisfied, then, given that prices are volatile, collusion would be hampered by the difficulty of distinguishing cheating from a conforming price set by a low marginal cost firm.

D.9.3 MSW's Supermarkup Model.

The conduct model is not a fully-specified model of collusion because it does not specify the incentives that cause firms to deviate from maximizing their own profits. Some collusion models cannot be tested using the conduct framework, but the MSW supermarkup model can be tested. MSW assume that, every fiscal year, both before and after the JV, a price leader suggests a "supermarkup" on top of Bertrand Nash prices that domestic brewers should charge. If a domestic firm fails to charge the supermarkup, a punishment phase ensues, but in a CI subgame perfect equilibrium, the suggested supermarkup will satisfy the incentive-compatibility constraints (ICCs). Prices may increase after a merger if the ICCs are relaxed. We can test this model by using an appropriately defined domestic product market-fiscal year fixed effect to control for the supermarkup. If the "supermarkup on Nash" theory is correct, estimates of conduct κ parameters should be equal to zero once the fixed effects are included. We explain this approach in more detail before presenting the results from two separate versions of the test.

MSW assume that each fiscal year, a price leader announces market-specific incentive-

[63]The rich fixed effects in columns (7)-(9) cause the ν_{imt}s to jump across fiscal years, so the estimated serial correlation falls.

compatible markups (m_{mt}), in dollars, above Nash prices that all domestic brewers should charge. Foreign brands are assumed to use static Nash pricing. This implies that, given m_{mt}, the first-order conditions for an AB product i are given in the following expression where $\widetilde{p^D} = p_{mt} - m_{mt}$ for domestic products and $\widetilde{p^I}(\widetilde{p^D})$ are Nash equilibrium prices of imported brands if domestic brewers charged $\widetilde{p^D}$:

$$p_{imt} - m_{mt} = W_{imt}\gamma + \frac{q_{imt}(\widetilde{p^D}, \widetilde{p^I}(\widetilde{p^D}))}{\frac{\partial q_{imt}}{\partial p_{imt}}(\widetilde{p^D}, \widetilde{p^I}(\widetilde{p^D}))} + \sum_{\substack{j \in AB \\ j \neq i}} \frac{\frac{\partial q_{jmt}}{\partial p_{imt}}(\widetilde{p^D}, \widetilde{p^I}(\widetilde{p^D}))}{\frac{\partial q_{imt}}{\partial p_{imt}}(\widetilde{p^D}, \widetilde{p^I}(\widetilde{p^D}))}(p_{jmt} - m_{mt} - c_{jmt}) + \nu_{imt}. \quad (6)$$

The first-order conditions for an imported product k (say a Heineken (H) product) are the standard static first-order conditions

$$p_{kmt} = W_{kmt}\gamma + \frac{q_{kmt}(p)}{\frac{\partial q_{imt}}{\partial p_{imt}}(p)} + \sum_{\substack{l \in H \\ l \neq k}} \frac{\frac{\partial q_{lmt}}{\partial p_{lmt}}(p)}{\frac{\partial q_{kmt}}{\partial p_{kmt}}(p)}(p_{lmt} - c_{lmt}) + \nu_{kmt}.$$

To test the model, we assume that the imported brands do use static best responses, and we test whether FOCs such as (6) describe the pricing of domestic producers. In particular we do this by generalizing the model to allow for a "conduct" parameter, i.e.,

$$p_{imt} - m_{mt} = W_{imt}\gamma + \frac{q_{imt}(\widetilde{p^D}, \widetilde{p^I}(\widetilde{p^D}))}{\frac{\partial q_{imt}}{\partial p_{imt}}(\widetilde{p^D}, \widetilde{p^I}(\widetilde{p^D})} + ...$$

$$\sum_{\substack{j \in AB \\ j \neq i}} \frac{\frac{\partial q_{jmt}}{\partial p_{imt}}(\widetilde{p^D}, \widetilde{p^I}(\widetilde{p^D}))}{\frac{\partial q_{imt}}{\partial p_{imt}}(\widetilde{p^D}, \widetilde{p^I}(\widetilde{p^D}))}(p_{jmt} - m_{mt} - c_{jmt}) + \kappa \sum_{k \in M,C} \frac{\frac{\partial q_{kmt}}{\partial p_{imt}}}{\frac{\partial q_{imt}}{\partial p_{imt}}}(p_{kmt} - c_{kmt}) + \nu_{imt}.$$

where, if the supermarkup explanation is correct, $\kappa = 0$. The intuition for the test is that if the supermarkup really is a constant markup on a static Nash price then, *controlling for supermarkup using an appropriately defined fixed effect*, price-setting should not be affected by the incremental effect that a price has on the profits of other domestic brewers. On the other hand, if there is an alternative type of deviation from CI Nash pricing then the estimated κ may be significantly different from zero.

Testing the Supermarkup Model: Linear Controls. We use two different implementations of the test. The first is easy to implement (which means that we can use it for monthly data) but relies on deviating from MSW's precise assumptions by assuming that supermarkups enter the FOCs linearly. Specifically, suppose that a domestic product i in market m has marginal cost c_{imt}, and that the collusive plan operates by each domestic product being priced according to static Nash best responses as if its marginal costs are $c_{imt} + m'_{mt}$ rather than just c_{imt}, where m'_{mt} is the supermarkup.[64] One interpretation would be that the domestic firms act as if they have to pay higher marginal retailing costs than they actually do, a form of tacit collusion that might be hard to detect. In this case, the first-order condition for an AB product is simply

$$p_{imt} = W_{imt}\gamma + m'_{mt} + \frac{q_{kmt}(p)}{\frac{\partial q_{imt}}{\partial p_{imt}}(p)} + \sum_{\substack{j \in AB \\ j \neq i}} \frac{\frac{\partial q_{jmt}}{\partial p_{imt}}(p)}{\frac{\partial q_{imt}}{\partial p_{imt}}(p)} (p_{jmt} - m'_{mt} - c_{jmt}) + \nu_{imt}. \tag{7}$$

and, when we generalize to allow for conduct coefficients, the estimating equation becomes

$$p_{mt} = W_{mt}\gamma + m_{mt} - \left(\Omega_{mt}(\kappa) \circ \left[\frac{\partial s_{mt}(p_t, \theta^D)}{\partial p_{mt}} \right] \right)^{-1} s_{mt}(p_{mt}) + \nu_{mt}. \tag{8}$$

This equation has the nice feature that the level of demand and the demand derivatives only depend on **observed** prices, and the unobserved supermarkup enters linearly. This theory can be tested by including domestic market-fiscal year fixed effects to control for m'_{mt}, and testing if conduct parameters equal zero. We use the domestic rival distance measures, their ξs (averaged across their portfolios either before or after the JV) and interactions of these variables as excluded instruments that identify the conduct parameters.

Columns (7)-(9) of Table D.5 present the results. As already discussed, we can reject $\kappa = 0$ before the JV at any significance level, and we cannot reject the hypothesis that "conduct" was the same before and after the JV.

Testing the Supermarkup Model: Nonlinear Controls. In the actual MSW model, supermarkups enter the first-order conditions non-linearly by affecting the values of the demand derivatives which, for domestic products, need to be evaluated at $\widetilde{p^D} = p_{mt} - m_{mt}$. The second verison of our test therefore estimates non-linear market-fiscal year fixed effects for domestic products, where, as these fixed effects are varied, we re-evaluate the demand derivative matrix

[64]The incentive-compatibility constraints would determine the magnitude of the m_{mt} supermarkup terms.

and resolve for the Nash prices that the imported brands would charge in response. This potentially creates a very large computational burden, especially when using the RCNL demand model, even if we use quarterly data. To make estimation feasible, we therefore proceed as follows.

First, we estimate all of the parameters, including the conduct parameters and the linear parameters, separately for each fiscal year, so that we are only estimating 40 (39 supermarkup fixed effects and 1 conduct parameter) nonlinear parameters at a time. We report the conduct coefficients for 2005/6, 2006/7, 2009/10, and 2010/11 fiscal years (i.e., two full fiscal years before the JV and after the JV), but we also estimate them for the partial fiscal years in the sample, and the estimated coefficients are similar, but less precise. We expect separate estimation to reduce the econometric efficiency and the power of our test, as will the fact that we do not restrict the supermarkups to be consistent with cross-market incentive compatibility constraints on the domestic brewers. However, in practice, our estimates of the conduct parameters are precise. We also use the quarterly supply and RCNL demand model. Recall that the results using this model in columns (3) and (6) of Table D.5 were the most favorable to the hypothesis of Nash pricing before the JV.

Second, and more importantly, rather than recomputing demand derivatives, import best responses prices and inverting matrices to back out implied marginal costs many hundreds of times during estimation, we use interpolation from values that are pre-computed. Specifically, before estimation, we compute implied marginal costs for each observed product-market-quarter observation on a grid of supermarkups ($m_{mt} = \{0, 0.25, 0.50, ..., 6\}$) and conduct parameters ($\kappa = \{0, 0.05, ..., 1.1\}$) then use cubic interpolation to get the required values during estimation (restricting the supermarkups and conduct parameters to lie within these ranges). As a result, the computational burden for each function evaluation involves the computation of around 6,000 cubic interpolations.

As usual, one might be skeptical about a researcher's ability to simultaneously estimate 40 nonlinear parameters. However, in practice, MATLAB's `fmincon` algorithm works very well on this problem even when it uses numerical derivatives, and it delivers the same estimates from alternative starting values. The conduct parameter estimates are also comfortingly consistent with those from testing version 1 of the supermarket model.

The results are reported in Column (10) of Table D.5. Consistent with column (9), the reported conduct parameters are precisely estimated and are between 0.9 and 1, and, because estimated supermarkups are also positive, most of the implied marginal costs are negative. Therefore, we can clearly reject the MSW formulation of CI collusion, although, as we have emphasized, this does not imply that all models of CI collusion can be rejected.