Oligopolistic Price Leadership and Mergers: The United States Beer Industry*

Nathan H. Miller[†]
Georgetown University

Gloria Sheu[‡]
Federal Reserve Board

Matthew C. Weinberg[§]
The Ohio State University

December 18, 2020

Abstract

We study a repeated game of price leadership in which one firm, the leader, proposes region-specific supermarkups over Bertrand prices to a coalition of rivals. The supermarkups and the firms' marginal costs can be recovered from scanner data using the structure of the model. In the beer industry, we find that price leadership increases profit relative to Bertrand by 17% over 2006-2007 and 22% over 2010-2011, with the change mostly due to consolidation. We use simulations to examine two mergers, and find that they relax incentive compatibility constraints and increase prices. These coordinated effects arise even with substantial efficiencies.

Keywords: price leadership, coordinated effects, mergers
JEL classification: K21; L13; L41; L66

*This material is based on work supported by the National Science Foundation under Grant Nos. 1824318 (Miller) and 1824332 (Weinberg). We thank many seminar and conference participants for insightful comments and suggestions. All estimates and analyses in this paper based on IRI data are by the authors and not by IRI. Much of the research was conducted while Gloria Sheu was a staff economist at the U.S. Department of Justice. The analysis and conclusions set forth are those of the authors and do not indicate concurrence by other members of the Board research staff or the Board of Governors. Furthermore, the views expressed here should not be purported to reflect those of the U.S. Department of Justice.

[†]Georgetown University, McDonough School of Business, 37th and O Streets NW, Washington DC 20057. Email: nhm27@georgetown.edu.

[‡]Board of Governors of the Federal Reserve System, 20th Street and Constitution Avenue NW, Washington DC 20551. Email: gloria.sheu@frb.gov.

[§]The Ohio State University, 410 Arps Hall, 1945 N. High Street, Columbus OH 43210. Email: weinberg.133@osu.edu.

1 Introduction

Firms in concentrated industries sometimes change their prices by similar magnitudes, with the changes initiated by a single firm. We follow Bain (1960) in referring to this pricing pattern as *oligopolistic price leadership*. The subject has a long history in economics. Anecdotal examples are discussed in Scherer (1980) and an older series of articles (e.g., Stigler (1947); Markham (1951); Oxenfeldt (1952)). More recent studies utilizing extremely detailed data document follow-the-leader pricing in retail industries ranging from supermarkets, pharmacies, and gasoline (Clark and Houde (2013); Seaton and Waterson (2013); Chilet (2018); Lemus and Luco (2018); Byrne and de Roos (2019).[1] However, as these studies are largely descriptive, existing research does not examine the effectiveness of price leadership in supporting supracompetitive markups, explore implications for welfare, nor provide a framework for the analysis of counterfactuals.[2]

This paper presents an empirical model of oligopolistic price leadership that can be evaluated with market level data on prices and quantities. In each period of an infinitely repeated game, the leader makes a non-binding price announcement and then all firms set prices simultaneously. The price announcement is cheap talk that shapes firm beliefs and facilitates supracompetitive pricing. We apply the model to a setting that exhibits such price leadership behavior—the beer industry of the United States. We recover the marginal costs and markups of each product from the data using first order conditions for profit maximization and the demand estimates of Miller and Weinberg (2017). A comparison to Bertrand equilibrium, obtained with a counterfactual simulation, allows us to quantify the implications of price leadership for firms and consumers.

Our modeling approach provides a framework for evaluating the coordinated effects of mergers in markets characterized by price leadership. To illustrate, we examine two mergers in particular, and show that they relax incentive compatibility (IC) constraints and increase prices. These results obtain even in the presence of marginal costs efficiencies sufficient to offset unilateral effects, which are the price changes under the common assumption of static price competition before and after the merger. Previously, the empirical industrial organization literature has provided little in the way of methodologies that could be used to

[1] See also the discussions in Lanzillotti (2017) and Harrington and Harker (2018). In the popular press, see "Drugmakers Find Competition Doesn't Keep a Lid on Prices" by Jonathan D. Rockoff, Wall Street Journal, November 27, 2016 and "Your Chocolate Addiction is Only Going to Get More (and More, and More) Expensive" by Roberto A. Ferdman, Washington Post, July 18, 2014.

[2] The study by Clark and Houde (2013) is an exception in that it uses a repeated pricing game to study the efficacy of a strategy employed by a known cartel of gasoline retailers.

guide coordinated effects analysis. Indeed, our research is among the first to formally model coordinated effects in real-world markets.[3]

We organize the paper as follows. We start in Section 2 with a description of U.S. brewing markets. In scanner data spanning 2001-2011, a handful of brewers account for the bulk of retail revenue. In the earlier years of the sample, these firms are Anheuser Busch (ABI), SABMiller, Molson Coors, Grupo Modelo, and Heineken. In the later years, SABMiller and Molson Coors are replaced with their joint venture, MillerCoors; we often refer to this consolidating event as the Miller/Coors merger. We summarize the qualitative evidence of price leadership behavior, citing in particular to legal documents filed by the Department of Justice (DOJ) alleging that ABI pre-announces its annual list price changes as a signal to competitors, and that MillerCoors tends to follow. We also discuss data sources, provide additional summary statistics and stylized facts, and describe the demand model of Miller and Weinberg (2017), which we take as given in this paper.

We then formalize the model of oligopolistic price leadership in Section 3. Firms compete across multiple geographic regions in an infinitely repeated game of perfect information. Each period has two stages. In the first, the leader announces non-binding and region-specific *supermarkups* above Bertrand prices. On the equilibrium path, a set of coalition firms, comprised of the leader and its followers, adopt the supermarkups in a subsequent pricing stage. The leader selects the supermarkups to maximize its profit, subject to the IC constraints of the followers and, in order for the announcement to be credible, itself. The leader also accounts for the reaction of fringe firms, which price to maximize static profit functions. Deviation, which occurs only off the equilibrium path, is punished with reversion to Bertrand pricing in all regions. These strategies constitute a subgame perfect equilibrium (SPE).

We discuss the empirical implementation in Section 4. Our approach relies on a pair of identification results: (i) marginal costs can be recovered, given any supermarkups, using first order conditions for static profit maximization, and (ii) the supermarkups can be recovered using another set of first order conditions that arises from the leader's constrained maximization problem. The latter result requires an *ex ante* assumption on a reduced-form timing parameter, which summarizes the patience of firms and various (unknown) aspects of the game, such as the duration of punishment. Larger timing parameters imply higher supermarkups and thus lower marginal costs for coalition firms, so with some prior knowledge

[3]We refer readers to Baker (2001, 2010) and Harrington (2013) for a summary of the legal literature on coordinated effects. The theoretical literature includes Compte et al. (2002); Vasconcelso (2005); Ivaldi et al. (2007); Bos and Harrington (2010); and Loertscher and Marx (2020). Empirical models include Davis and Huse (2010) and Igami and Sugaya (2020).

of costs it is possible to evaluate the timing assumptions *ex post*. We use an orthogonality condition for this purpose, namely an assumption that ABI's marginal costs do not change differently from those of Modelo and Heineken, on average, with the Miller/Coors merger.

We then summarize the empirical results and analyze equilibrium in Section 5. Using the timing parameter that best satisfies the identifying assumption, we recover average supermarkups of \$1.20 in fiscal year 2007, just before Miller/Coors, and \$1.80 in fiscal year 2010, just after. The change in supermarkups between 2007 and 2010 reflects that the merger created slack in the binding IC constraint and also greater symmetry among coalition firms, though new cost and demand conditions also contribute. The difference between industry profits under price leadership and profits under static Bertrand competition is 17% and 22% of Bertrand profits in 2007 and 2010, respectively. The reduction in consumer surplus is 154% and 170% of the producer surplus gain in those two years.

Supermarkups tend to be higher in regions where ABI has large market shares, and lower in regions where Coors (in 2007) and MillerCoors (in 2010) have large market shares. This reflects the Kuhn-Tucker conditions that characterize the solution to ABI's constrained maximization problem: ABI benefits more from a higher supermarkup if it has a large market share, and the effect of a higher supermarkup on the binding IC constraint is greater if Coors and MillerCoors have large market shares. Relatedly, we use counterfactual simulation to explore the role of multi-market (here, multi-region) contact. The results indicate that multi-market contact affects the spatial dispersion of supermarkups, but less so after the Miller/Coors merger due to the enhanced symmetry among coalition firms.

In Section 6, we use the model to examine the coordinated effects of the Miller/Coors merger and ABI's proposed acquisition of Modelo, which was approved in 2013 by the DOJ only after the Modelo brands were sold to a third party. We model the latter merger as it would have occurred without the divestiture. In both instances we find that the merger loosens the IC constraint of the binding firm, resulting in an increase for the supermarkup on domestic beers of \$0.50 due to Miller/Coors and an increase of \$0.40 due to ABI/Modelo. Miller/Coors combines the two smaller firms from the pricing coalition into one firm that then faces demand and cost conditions that are more similar to the leader, ABI. The ABI/Modelo merger, as originally proposed, brings the largest outside firm into the coalition, which lessens competition from the pricing fringe. In neither case are the price effects mitigated by marginal cost efficiencies.

We conclude in Section 7 with a short summary and a discussion of some of the more important modeling assumptions, with an eye toward informing future research efforts. The Appendix includes additional details on the data, a set of theoretical results and proofs, a description of computational methods, and assorted additional analyses.

1.1 Literature Review

Our research is methodologically most similar to Igami and Sugaya (2020), which studies the vitamin C cartel of the 1990s. Among the main findings of that paper is that unexpected shocks to demand and fringe supply undermined incentive compatibility and led to the collapse of the cartel. As in our research, Igami and Sugaya estimate the structural parameters of a supergame in which trigger strategies sustain supracompetitive prices, and rely on counterfactual simulations to recover the profit terms that enter the IC constraints. There are also notable differences in the models. For example, Igami and Sugaya assume all firms engage in maximal collusion or revert to Cournot equilibrium forevermore. Some interesting aspects of our model, such as partial coalitions, multi-market contact, and the leader's ability to adjust the prices to satisfy IC constraints, are not present.

Eizenberg et al. (2020) and Eizenberg and Shilian (2019) also estimate IC constraints in empirical settings. The first of these estimates demand for hummus salad and instant coffee in the Israeli grocery sector, and recovers marginal costs with an assumption of Bertrand competition. It then evaluates hypothetical coordination and determines that multi-market contact would not substantially relax IC constraints, a result that is attributed to symmetry across the two product categories. In our setting, the degree of symmetry also affects the impact of multi-market contact. The second paper estimates conduct in 40 food categories and imputes the minimum discount factor necessary to support the conduct in SPE.

More broadly, our research relates to articles that seek to understand the equilibrium concept that governs competition in specific markets. Two of the more prominent papers focus on Bertrand equilibrium and joint profit maximization (Bresnahan (1987); Nevo (2001)), though Stackleberg leadership and various other possibilities also have been examined (e.g., Gasmi et al. (1992); Slade (2004); Rojas (2008)). Other studies use a conduct parameter approach to identify changes in the intensity of competition, without taking a stance on the precise equilibrium concepts (e.g., Porter (1983); Igami (2015); Miller and Weinberg (2017); Michel and Weiergraeber (2018)). Finally, there is an empirical literature that tests whether multi-market contact leads to higher prices, without modeling the underlying game (e.g., Evans and Kessides (1994); Ciliberto and Williams (2014); Shim and Khwaja (2017)).

The price leadership model itself draws on a number of theoretical contributions. An important precursor is the canonical Rotemberg and Saloner (1986) model of collusion, in which there is perfect information and collusive prices adjust so that deviation does not occur along the equilibrium path.[4] Our treatment of multi-market contact extends a model of differentiated-products price competition developed in Bernheim and Whinston (1990). We develop the connections to the theoretical literature more extensively in Section 3, after presenting the model of price leadership.

2 The U.S Beer Market

2.1 Background

Beer is differentiated along multiple dimensions, including taste, calories, brand image, and package size. Most beer sold in the Unites States is lager style, and a handful of brewers dominate this market. This is despite recent growth of craft brewers, which tend to specialize in ales. Retailers and distributors also play a role in the supply chain, but in a different way than most industries because of regulation on the sales and distribution of alcohol dating back to prohibition. Large brewers are prohibited from selling beer directly to retail outlets. Instead, they typically sell to state-licensed distributors, who, in turn, sell to retailers. Payments along the supply chain cannot include slotting fees, slotting allowances, or other fixed payments between firms.[5] While retail price maintenance is technically illegal in many states, in practice, distributors are often induced to sell at wholesale prices set by brewers (Asker (2016)).

Table 1 summarizes the revenue shares of the major brewers over 2001-2011. In the early years of the sample, Anheuser-Busch, SABMiller, and Molson Coors (domestic brewers) account for 61-69% of revenue while Grupo Modelo and Heineken (importers) account for another 12-16% of revenue.[6] Midway through the sample, in June 2008, SABMiller and Molson Coors consolidated their U.S. operations into the MillerCoors joint venture. The DOJ elected not to challenge consolidation on the basis that cost savings in distribution likely would offset any loss of competition.[7] In the same year, Inbev purchased Anheuser-Busch (forming ABI). As Inbev previously sold brands with limited sales in the US, such as

[4] A repeated game in which oligopolistic price leadership emerges is provided in Rotemberg and Saloner (1990). As their model incorporates asymmetric information, price announcements have informational and strategic content. Our model does not include asymmetric information.

[5] The relevant statutes are the Alcoholic Beverage Control Act and the Federal Alcohol Administration Act, both of which are administered by the Bureau of Alcohol, Tobacco and Firearms (see their 2002 advisory at `https://www.abc.ca.gov/trade/Advisory-SlottingFees.htm`, last accessed November 4, 2014).

[6] We refer to the first three firms as "domestic" because their beer is brewed in the United States.

[7] Subsequent academic research suggests that sizable costs savings were realized but were dominated by adverse competitive effects (Ashenfelter et al. (2015), Miller and Weinberg (2017)).

Table 1: Revenue-Based Market Shares

Year	ABI	MillerCoors	Miller	Coors	Modelo	Heineken	Total
2001	0.37	.	0.20	0.12	0.08	0.04	0.81
2003	0.39	.	0.19	0.11	0.08	0.05	0.82
2005	0.36	.	0.19	0.11	0.09	0.05	0.79
2007	0.35	.	0.18	0.11	0.10	0.06	0.80
2009	0.37	0.29	.	.	0.09	0.05	0.80
2011	0.35	0.28	.	.	0.09	0.07	0.79

Notes: The table provides revenue shares over 2001-2011. Firm-specific revenue shares are provided for ABI, Miller, Coors, Modelo, and Heineken. The total across these firms also is provided. The revenue shares incorporate changes in brand ownership during the sample period, including the merger of Anheuser-Busch (AB) and Inbev to form A-B Inbev (ABI), which closed in April 2009, and the acquisition by Heineken of the FEMSA brands in April 2010. All statistics are based on supermarket sales recorded in IRI scanner data.

Stella Artois, the transaction had only minor consequences for market structure.

There have been other notable transactions after the sample period. In 2013, ABI acquired Grupo Modelo. The DOJ obtained a settlement under which the rights to the Grupo Modelo brands in the U.S. transferred to Constellation, at that time a major distributor of wine and liquor.[8] In 2016, ABI acquired SABMiller. In order to gain DOJ approval, SABMiller sold its stake in the MillerCoors joint venture to Molson Coors. Finally, ABI has purchased a number of craft breweries over the last decade, and now owns Goose Island, Devil's Backbone, and Kona, among a number of other craft brands.

2.2 Price Leadership in the Beer Industry

There is extensive qualitative evidence that competition among brewers involves price leadership behavior. Legal documents filed in 2013 by the DOJ to enjoin the ABI/Modelo acquisition allege that:

> ABI and MillerCoors typically announce annual price increases in late summer for execution in early fall. In most local markets, ABI is the market share leader and issues its price announcement first, purposely making its price increases transparent to the market so its competitors will get in line. In the past several years, MillerCoors has followed ABI's price increases to a significant degree.[9]

[8]The press release of the DOJ provides details on the settlement. See `https://www.justice.gov/opa/pr/justice-department-reaches-settlement-anheuser-busch-inbev-and-grupo-modelo-beer-case`, last accessed February 13, 2019.

[9]Para 44 of the Complaint in *US v. Anheuser-Busch InBev SA/NV and Grupo Modelo S.A.B. de C.V.*

The legal documents do not specify whether these pricing practices were used prior to the Miller/Coors merger in 2008. However, two prominent industry studies describe price leadership as occurring throughout the latter half of the twentieth century (Greer (1998), Tremblay and Tremblay (2005)).[10] Further, a recent enforcement action of the DOJ, related to ABI's acquisition of the Craft Brewers Alliance (CBA), suggests that price coordination is ongoing.[11] We interpret these descriptions as suggesting that price leadership occurs throughout the sample period, and maintain that assumption in our empirical analysis.[12]

The qualitative evidence guides other modeling assumptions that we make as well. First, price leadership behavior in the industry does not appear to involve Modelo or Heineken. The legal filings state that Modelo adopted a "Momentum Plan" to "grow Modelo's market share by shrinking the price gaps."[13] Drennan et al. (2013), an article written by DOJ economists, notes that "[i]n internal strategy documents, ABI has repeatedly complained about pressure resulting from price competition with Modelo brands."[14]

Second, trigger strategies may be important in sustaining supracompetitive pricing. Tremblay and Tremblay (2005, p. 168) state "Anheuser-Busch has used and threatened to use substantial price reductions to punish rivals...."[15] Price wars appear to be relatively infrequent, however, as only one example is provided (occurring over 1953-1955).

Third, the legal documents filed by the DOJ in 2013 provide some support for our assumption that the leader's price announcement serves as an equilibrium selection device. The following passage quotes from the business documents of ABI:

> ABI's Conduct Plan emphasizes the importance of being "Transparent – so competitors can clearly see the plan;" "Simple – so competitors can understand the

[10]For example, Tremblay and Tremblay (2005, p. 49) states that "Anheuser-Busch serves as the price leader for the industry" and that "Most other brewers... key to Budweiser, usually matching Bud's price for premiums and going somewhat above it or below it for superpremiums and populars." Tremblay and Tremblay (2005, p. 168) states that the purpose of leadership is to "maintain high prices."

[11]The DOJ press release describes the acquisition and divestiture, and states that "[b]y eliminating CBA's Kona brand as a competitive restraint, ABI would also likely have greater ability to facilitate price coordination, resulting in higher prices...." See `https://www.justice.gov/opa/pr/justice-department-requires-divestiture-order-anheuser-busch-acquire-craft-brew-alliance`, last accessed September 19, 2020.

[12]Miller and Weinberg (2017) cite to evidence culled from the annual reports of ABI and SABMiller that suggest competition was relatively tough over 2005-2008. This is consistent with our results, which indicate that the Miller/Coors merger resulted in higher supermarkups.

[13]Para 49 of the Complaint in *US v. Anheuser-Busch InBev SA/NV and Grupo Modelo S.A.B. de C.V.*

[14]Drennan et al. (2013), p., 295. The legal filings also speak to this. For example, the Competitive Impact Statement (p. 8) states that "[b]y compressing the price gap between high-end and premium brands, Modelo's actions have increasingly limited ABI's ability to lead beer prices higher, resulting in higher prices...." The legal filings do not address Heineken specifically, though their prices are similar to Modelo's in the data.

[15]See also Greer (1998, p.50): "Anheuser-Busch's strategy includes cutting price to discipline rivals...."

> plan;" "Consistent – so competitors can predict the plan;" and "Targeted – consider competition's structure." By pursuing these goals, ABI seeks to "dictate consistent and transparent competitive response."[16]

We view this passage as suggesting that the primary purpose of ABI's price announcements is to provide strategic clarity for MillerCoors. If this interpretation is correct then there is a tight connection between price announcements in the beer industry and in our model.

2.3 Data and Prices

We use retail scanner data from the IRI Academic Database (Bronnenberg et al. (2008)), which contains weekly revenue and unit sales by UPC code for a sample of stores over 2001-2011. We restrict attention to supermarkets, which accounted for 26% of off-premise beer sales in 2011 (McClain (2012)).[17] We aggregate the UPCs to the brand×size level. For convenience, we often refer to brand×size combinations as "products." We focus on 13 flagship brands sold as six packs, 12 packs, and 24 packs.[18] We measure quantities based on 144-ounce equivalent units, the size of a 12 pack, and measure price as the ratio of revenue to equivalent unit sales. These choices comport with Miller and Weinberg (2017).

In our main supply-side analysis, we aggregate the weeks to quarters and use data from 37 distinct geographic regions. Thus, the primary unit of observation is a product-region-quarter. Because list price adjustments take effect in the fall (see the previous section), we further group quarters into "fiscal years" that begin in October and end in the following September. Thus, for example, fiscal year 2007 comprises to the fourth quarter of 2006 and the first three quarters of 2007. We restrict attention to the fiscal years 2006, 2007, 2010, and 2011, each of which is fully contained in the sample period. We exclude earlier years because our demand model relies on household demographics from the Public Use Microdata Sample (PUMS) of the American Community Survey, which is available starting in 2005. We omit fiscal years 2008 and 2009 due to their proximity to the Miller/Coors merger, and fiscal year 2005 because it is only partially covered in the data.[19]

[16] Para 46 of the Complaint in *US v. Anheuser-Busch InBev SA/NV and Grupo Modelo S.A.B. de C.V.*

[17] The other major sources of off-premise beer sales are liquor stores (38%), convenience stores (26%), mass retailers (6%), and drugstores (3%). The price and quantity patterns that we observe for supermarkets also exist for drug stores, which are in the IRI Academic Database. See Miller and Weinberg (2017).

[18] We combine 24 packs and 30 packs in the construction of these products because whether 24 or 30 packs are sold tends to depend on region-specific historical considerations. We exclude 18-packs and promotional package sizes, which generate fewer sales.

[19] We focus on quarterly aggregation because it is computationally less burdensome than monthly aggregation. However, the empirical variation we exploit is similar to what is used in the simpler model of Miller and Weinberg (2017), and those results are robust to monthly and quarterly aggregation.

We rely on a number of other sources to complete the data set. We use Google Maps to obtain the driving miles between each IRI region and the nearest brewery for each of the domestic products. For the imported brands, we obtain the driving miles between the regions and the nearest port into which the beer is shipped.[20] Our measure of "distance" is the multiplicative product of driving miles and diesel fuel prices, which we obtain from the Energy Information Agency of the Department of Energy. This allows us to capture variation in transportation costs that arises both cross-sectionally, based on the location of regions and breweries, and inter-temporally, based on fluctuations in fuel costs. All prices and incomes are deflated using the CPI and are reported in 2010 dollars. See Appendix A for additional details about the data.

Table 2 shows the average price of each product (brand×size) in fiscal year 2011, along with the share of volume among the products in the sample. Domestic brands and larger package sizes tend to have lower prices. Volume shares increase in package sizes for the domestic brands, whereas 12 packs have the greatest volume shares for the imported brands. The most popular domestic brands are Bud Lite, Coors Light, Miller Lite, and Budweiser. The most popular imported brands are Corona Extra and Heineken.

Figure 1 shows the time path of average retail prices over 2001-2011 for each firm's most popular 12 pack. The red vertical line at June 2008 marks the Miller/Coors merger. As shown, the prices of domestic beers increase abruptly after the merger, while import prices continue on trend. Notably, the price increases of ABI are commensurate with those of MillerCoors. Miller and Weinberg (2017) determines that the data are difficult to explain as a shift among Bertrand equilibria, absent a sizable ABI marginal cost increase. These price data also are a focus of the present study, and we show that they can be rationalized with a price leadership framework featuring binding IC constraints.

2.4 Demand

We rely on the random coefficient nested logit (RCNL) model of Miller and Weinberg (2017) to characterize consumer demand. Our preferred specifications are RCNL-1 and RCNL-2, which allow income to affect the price parameter. This incorporates consumer heterogeneity in the willingness-to-pay for more expensive imported beers and smaller package sizes.

[20] We obtain the location of Heineken's primary ports from the website of BDP, a logistics firm hired by Heineken to improve its operational efficiency. See `http://www.bdpinternational.com/clients/heineken/`, last accessed on February 26, 2015. The ports include Baltimore, Charleston, Houston, Port of Long Beach, Miami, Seattle, Oakland, Boston, and New York. We measure the shipping distance for Grupo Modelo brands as the driving distance from each retail location to Ciudad Obregon, Mexico.

Table 2: Prices and Volume Shares in Fiscal Year 2011

Brand	Brewer	6 Packs Share	6 Packs Price	12 Packs Share	12 Packs Price	24 Packs Share	24 Packs Price	All Share
Bud Light	ABI	0.020	11.61	0.074	10.02	0.193	8.16	0.288
Budweiser	ABI	0.011	11.60	0.031	10.01	0.074	8.15	0.116
Coors	MillerCoors	0.001	11.75	0.004	10.10	0.011	8.07	0.017
Coors Light	MillerCoors	0.010	11.63	0.042	10.08	0.107	8.12	0.159
Corona Extra	Modelo	0.010	15.85	0.040	13.02	0.018	12.56	0.067
Corona Light	Modelo	0.005	15.89	0.019	13.07	0.002	12.62	0.027
Heineken	Heineken	0.007	16.22	0.030	13.35	0.007	12.80	0.044
Heineken Light	Heineken	0.002	16.33	0.007	13.42	0.001	11.79	0.010
Michelob	ABI	0.002	12.43	0.005	10.81	0.004	8.03	0.011
Michelob Light	ABI	0.008	12.56	0.025	10.88	0.017	8.63	0.049
Miller Genuine Draft	MillerCoors	0.003	11.65	0.007	10.05	0.012	8.14	0.022
Miller High Life	MillerCoors	0.004	9.15	0.022	7.92	0.032	6.69	0.058
Miller Lite	MillerCoors	0.009	11.60	0.046	10.09	0.110	8.13	0.164

Notes: This table provides the volume share and average price for each brand–size combination in the fiscal year 2011. The volume shares are among the brands shown, and so sum to one. Prices are per 144 ounces (the size of a 12 pack).

Among these two specifications, we focus on RCNL-2, so as to comport with our quarterly observations. The median product-level demand elasticity is -4.74 whereas the median market demand elasticity is -0.60, indicating that consumers tend to substitute among beer products, rather than between beer and the outside good.[21] Miller and Weinberg (2017) find these demand elasticities to be reasonably robust across a range of specification choices. We describe the RCNL demand model in greater detail in Appendix E.

3 Model

3.1 Overview

We now develop the model of oligopoly price leadership. Let there be $i = 1, \ldots, F$ firms and $j = 1, \ldots, J$ differentiated, substitute products. Each firm i produces a subset $\mathbb{J}_i$ of all products. We take as given the existence of a pricing *coalition*, a set $\mathbb{C}$ which comprises a leader and at least one follower. Firms not in the coalition are considered part of the *fringe*. The allocation of firms to the coalition and fringe is fixed and predetermined. There are $r = 1, \ldots, R$ distinct geographic regions. In each period ($t = 1, 2, \ldots$) an economic state of demand and cost conditions is realized, and competition then plays out in two stages:

[21] Median firm-level elasticities are around three for the domestic brewers and five for the import brewers.

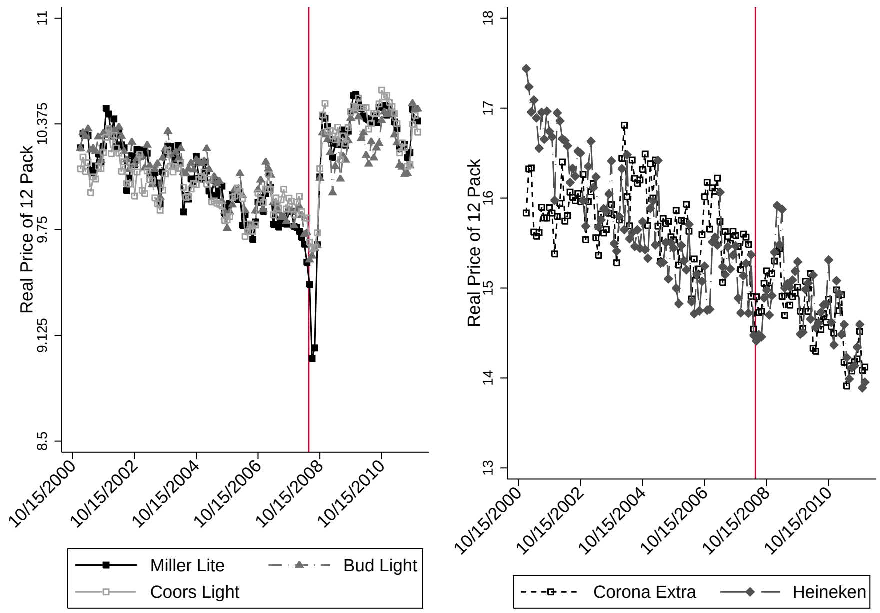

Figure 1: Average Retail Prices of Flagship Brand 12-Packs

Notes: The figure plots the national average price of a 12-pack over 2001-2011, separately for Bud Light, Miller Lite, Coors Light, Corona Extra and Heineken. The vertical axis is the natural log of the price in real 2010 dollars. The vertical bar drawn at June 2008 signifies the consummation of the Miller/Coors merger. Reproduced from Miller and Weinberg (2017).

(i) The leader announces non-binding supermarkups, $m_{rt} \geq 0$ for each region.

(ii) All firms price simultaneously, given $m_t = (m_{1t}, m_{2t}, \ldots, m_{Rt})$ and the history.

The game ends with probability $(1 - \phi)$ after each period. The non-binding announcements are cheap talk that shape beliefs in the pricing stage. Thus, they are not a theoretical necessity and could be replaced with an assumption on equilibrium selection.[22]

Firms maximize the present value of profit. Let $p_{rt} = (p_{1rt}, p_{2rt}, \ldots, p_{Jrt})$ be the prices in region r and period t, and let Ψ_t denote the economic state. The profit function of firm i is

$$\pi_{irt}(p_{rt}; \Psi_t) = \sum_{j \in \mathbb{J}_i} (p_{jrt} - mc_{jrt}(\Psi_t)) q_{jrt}(p_{rt}; \Psi_t) \tag{1}$$

where mc_{jr} is a constant marginal cost function and $q_{jr}(\cdot)$ is a differentiable demand function. Firms apply a discount factor, $\delta \in (0, 1)$, to calculate present values. We assume that the

[22] We assume that the supermarkup applies equally to all coalition products in its region for simplicity and computational tractability. In the empirical application, we consider a robustness exercise in which different supermarkups apply to products of different package sizes (Appendix D).

economic state is common knowledge and that firms observe all previous states, prices, quantities, and supermarkup announcements. These elements constitute the history that is considered in stage (ii) of each period. Firm actions do not affect the economic state. The equilibrium concept is subgame perfection.

We turn next to the building blocks of the price leadership model: static first order conditions, slack functions, and the leader's maximization problem. We then provide the strategies which constitute the *price leadership equilibrium* (PLE). Finally, we relate our model to the theoretical literature.

3.2 Static First Order Conditions

Differentiating equation (1) with respect to price obtains static first order conditions for (stage-game) profit maximization. Let $(p_{irt}, q_{irt}, mc_{irt})$ be vectors of firm i's prices, quantities, and costs in region r and period t, and let p_{-irt} contain the prices of competitors. Then the static first order conditions of firm i take the form

$$f(p_{irt}, p_{-irt}; \Psi_t) \equiv p_{irt} + \left[\frac{\partial q_{irt}(p_{irt}, p_{-irt}; \Psi_t)}{\partial p_{irt}}^T\right]^{-1} q_{irt}(p_{irt}, p_{-irt}; \Psi_t) - mc_{irt}(\Psi_t) = 0 \quad (2)$$

We assume there exists a unique $p^*_{irt}(p_{-irt}; \Psi_t)$ that solves this system of equations, for each firm i and any competitor prices. This assumption can be verified for the special case of logit demand by adapting an argument of Nocke and Schutz (2018). A number of coding checks suggest existence and uniqueness in our empirical application, but with the RCNL demand system this is not guaranteed. If *all* firms solve their static first order conditions then Bertrand equilibrium obtains, with $p^*_{irt}(p^*_{-irt}; \Psi_t)$ for all i. We collect the Bertrand prices in the vector p^B_{rt} for notational convenience.

Differentiated-products Bertrand pricing often is assumed in the empirical literature. If demand is known and the data contain Bertrand prices and quantities, then equation (2) identifies marginal costs (Rosse (1970)). In our price leadership model, the static first order conditions also have empirical content. Take as given that, along the equilibrium path, coalition firms set prices according to $p^{PL}_{irt}(m_{rt}) = p^B_{irt} + m_{rt}$, and fringe firms solve equation (2) holding fixed the coalition prices. Then marginal costs can be recovered for any given m_{rt}. We formalize the result with the following proposition:

Proposition 1 (Identification of Marginal Costs). *If the econometrician has knowledge of price leadership prices, the demand system, the identities of the coalition firms, and the* *supermarkup, then Bertrand prices and marginal costs are identified.*

Proof: The proof is constructive and proceeds in four steps, each of which is easily verified given the maintained assumptions. We enumerate the steps here as they are central to the empirical implementation:

1. Infer mc_{jrt} for the products of fringe firms from equation (2). This can be done because fringe firms maximize stage-game profit.

2. Obtain $p^B_{jrt} = p^{PL}_{jrt}(m_{rt}) - m_{rt}$ for the products of coalition firms.

3. Compute p^B_{jrt} for the products of fringe firms by simultaneously solving the best response function, given the marginal costs inferred from step 1 and holding the prices of coalition firms fixed at the Bertrand levels obtained from step 2.

4. Infer mc_{jrt} for the products of coalition firms from equation (2), evaluated at the Bertrand prices p^B_{rt} obtained in steps 2 and 3.

QED

3.3 Slack Functions

We now define the *slack function* of each firm, which provides the net present value of price leadership less that of deviation, given a set of supermarkups. We use the slack function to characterize the IC constraints in the leader's maximization problem. For present purposes, we assume that deviation profit is earned for $\tau_1 \geq 1$ periods, and that punishment takes the form of Bertrand pricing for $\tau_2 \geq 1$ periods. We verify later that these timing assumptions are consistent with the strategies that characterize the PLE.[23]

Some additional notation is necessary. Let the profit that firm i receives with price leadership be $\pi^{PL}_{irt}(m_{rt}; \Psi_t) \equiv \pi_{irt}(p^{PL}_{rt}(m_{rt}); \Psi_t)$, where the prices in $p^{PL}_{rt}(m_{rt})$ are as defined in the previous subsection. Let the deviation prices of firm i be those which arise if firm i solves its static first order conditions (equation (2)) and other firms set prices according to $p^{PL}_{rt}(m_{rt})$. Denoting these prices as $p^{D,i}_{rt}(m_{rt})$, we have deviation profit of $\pi^{D,i}_{irt}(m_{rt}; \Psi_t) \equiv \pi_{irt}(p^{D,i}_{rt}(m_{rt}); \Psi_t)$. Finally, let profit in Bertrand equilibrium be $\pi^B_{irt}(\Psi_t) \equiv \pi_{irt}(p^B_{rt}; \Psi_t)$.

[23] This rules out optimal punishments (e.g., Abreu (1986)). Even with explicit collusion, often cartels do not employ complex punishments, other than making transfers (Harrington and Skrzypacz (2011)).

The contribution of region r to the slack function of firm i is given by

$$\begin{aligned}\tilde{g}_{irt}(m_{rt};\delta,\phi,\tau_1,\tau_2,\tilde{m}_{rt},\Psi_t) \equiv & \pi^{PL}_{irt}(m_{rt};\Psi_t) + \frac{\phi\delta}{1-\phi\delta}\pi^{PL}_{irt}(\tilde{m}_{rt};\Psi_t) \\ & - \Bigg(\sum_{s=0}^{\tau_1-1}(\phi\delta)^s \pi^{D,i}_{irt}(m_{rt};\Psi_t) + \sum_{s=\tau_1}^{\tau_1+\tau_2-1}(\phi\delta)^s\pi^{B}_{irt}(\Psi_t) \\ & + \sum_{s=\tau_1+\tau_2}^{\infty}(\phi\delta)^s\pi^{PL}_{irt}(\tilde{m}_{rt};\Psi_t)\Bigg)\end{aligned} \tag{3}$$

where $\tilde{m}$ refers to the supermarkup expected in future periods involving price leadership. The first line is the present value of price leadership in the region, and the second line subtracts the present value of deviation followed by punishment and an (eventual) return to price leadership. Summing across regions, the slack function is given by

$$\tilde{g}_{it}(m_t;\delta,\phi,\tau_1,\tau_2,\tilde{m}_t,\Psi_t) \equiv \sum_{r=1}^{R}\tilde{g}_{irt}(m_{rt};\delta,\phi,\tau_1,\tau_2,\tilde{m}_{rt},\Psi_t) \tag{4}$$

for the vectors of supermarkups $m_t = (m_{1t}, m_{2t}, \dots, m_{Rt})$ and expected supermarkups $\tilde{m}_t = (\tilde{m}_{1t}, \tilde{m}_{2t}, \dots, \tilde{m}_{Rt})$. The slack function is positive if the present value of price leadership exceeds the present value of deviation, and negative otherwise.

Using the slack function directly in empirical work presents identification challenges. Suppose that the profit terms are known and that $\tilde{g}_{it}(m_t;\cdot) = 0$ for some coalition firm i. As the discount factor δ and the continuation probabilityϕ enter as multiplicative factors, knowledge that $\tilde{g}_{it}(m_t;\cdot) = 0$ is insufficient to disentangle the two parameters. Further, it can be verified that for any (τ_1,τ_2) there exists some $\phi\delta$ that satisfies the equality, meaning that $\phi\delta$ cannot be separately identified from (τ_1,τ_2) on the basis of $\tilde{g}_{it}(m_t;\cdot) = 0$ alone. Complicating matters is that deviation and punishment do not occur on the equilibrium path, a result that we develop shortly, so τ_1 and τ_2 cannot be discerned from data on equilibrium outcomes. Absent other evidence about the strategies played by firms off the equilibrium path, $\tilde{g}_{it}(m_t;\cdot) = 0$ provides only joint identification of $(\delta,\phi,\tau_1,\tau_2)$.

To help facilitate empirical progress, we construct an *equivalent slack function* in which the inter-temporal trade off is governed by a single reduced-form timing parameter. The equivalent slack function is

$$g_{it}(m_t;\eta,\tilde{m}_t,\Psi_t) \equiv \sum_{r=1}^{R} g_{irt}(m_{rt};\eta,\tilde{m}_{rt},\Psi_t) \tag{5}$$

where

$$g_{irt}(m_{rt};\eta,\tilde{m}_{rt},\Psi_t) \equiv \pi_{irt}^{PL}(m_{rt};\Psi_t) + \frac{\eta}{1-\eta}\pi_{irt}^{PL}(\tilde{m}_{rt};\Psi_t) - \left(\pi_{irt}^{D,i}(m_{rt};\Psi_t) + \frac{\eta}{1-\eta}\pi_{irt}^{B}(\Psi_t)\right) \tag{6}$$

and

$$\eta \equiv \eta(\delta,\phi,\tau_1,\tau_2) = \frac{(\phi\delta)^{\tau_1} - (\phi\delta)^{\tau_1+\tau_2}}{1-(\phi\delta)^{\tau_1+\tau_2}} \tag{7}$$

The equivalent slack function provides the net present value of price leadership less that of deviation in an infinitely-repeated version of the game featuring a single period of deviation profit and Grim Trigger punishment strategies.

We now clarify our notion of equivalence formally:

Proposition 2 (Equivalent Slack Function): *The slack function and the equivalent slack function are related according to*

$$\tilde{g}_{it}(m_t;\delta,\phi,\tau_1,\tau_2,\tilde{m}_t,\Psi_t) = \psi g_{it}(m_{rt};\eta,\tilde{m}_t,\Psi_t)$$

for some $\psi \in (0,1]$. Further, if $\tau_1 = 1$ then $\psi = 1$.

Proof: See Appendix B.1.

The two versions of the slack function both provide valid characterizations of the IC constraints because they share the same sign for any vector of supermarkups. Further, the timing parameter that solves the equation $g_{it}(m_t;\eta,\tilde{m}_t,\Psi_t) = 0$ summarizes the patience of firms, the continuation probability, and the durations of deviation and punishment that would solve the equation $\tilde{g}_{it}(m_t;\delta,\phi,\tau_1,\tau_2,\tilde{m}_t,\Psi_t) = 0$.

Two observations are appropriate at this point. First, our construction of the slack functions assumes that firms do not anticipate changes in the economic state. The assumption appears necessary for empirical work because the main alternatives are untenable—it is not clear how to model the processes that govern costs and demand, and a purely empirical approach is impossible because an infinite series of data would be required.[24] A consequence

[24]Igami and Sugaya (2020) maintains the same assumption. In terms of realism, some of the big changes in the sample, related to the Great Recession and gasoline prices for example, would have been difficult for firms to anticipate. On the other hand, the brewers probably had a decent sense about longer-term trends, the most important being shifting consumer preferences for craft beer. If information were available regarding

is that the model does not generate counter-cyclical pricing à la Rotemberg and Saloner (1986), which arises because the present value of deviation is smaller if current demand is weak relative to future demand. However, we can show that if the profit function grows at a constant rate then the equivalent slack function obtains nonetheless.

The second observation is that the slack functions allow the leader's announced supermarkups (m_t) to differ from expectations ($\tilde{m}_t$) about what the leader will announce in the future, with the same economic state. Firms' expectations play an important role in the slack functions and, ultimately, the IC constraints. For example, if firms expect supermarkups of zero in the future, then they also expect punishment for deviation to be inconsequential, and it follows that positive supermarkups cannot be sustained.[25] To make progress, we assume that expectations are rational in a sense that we formalize in the next subsection.

3.4 The Leader's Problem

We assume the leader announces supermarkups that maximize its profit subject to IC constraints. Without loss of generality, let firm 1 be the leader. Each period the leader chooses markups for each region, $m_t^* = (m_{1t}^*, m_{2t}^*, \ldots, m_{Rt}^*)$, given that each firms' beliefs about future markups are fixed. The solution is then given by

$$m_t^*(\tilde{m}_t, \eta, \Psi_t) = \arg\max_{m_t \geq 0} \sum_{r=1}^{R} \pi_{1rt}^{PL}(m_{rt}; \Psi_t) \quad s.t. \quad g_{it}(m_t; \eta, \tilde{m}_t, \Psi_t) \geq 0 \quad \forall i \tag{8}$$

We require rational expectations, so that firms' expectations for future supermarkups align with the solution to the leader's constrained maximization problem given those expectations (i.e., $m_t^*(\tilde{m}_t, \eta, \Psi_t) = \tilde{m}_t$). We assume the existence of a unique non-degenerate rational expectations solution.[26] We drop $\tilde{m}_t$ as a function argument henceforth, as we have $m_t^* = \tilde{m}_t$. For empirical purposes, it is helpful to derive the first order conditions that characterize the rational expectations solution. This can be done under two mutually exclusive assumptions:

Assumption 1: *IC does not constrain the leader, i.e.,* $g_{it}(m_t^*; \eta, \Psi_t) > 0 \ \forall i$.

Assumption 2: *IC constrains the leader, i.e.,* $g_{kt}(m_t^*; \eta, \Psi_t) = 0$ *for some* $k \in \mathbb{C}$.

the lengths of deviation and punishment—for example if documentary sources reveal firm beliefs—then the purely empirical approach could be implemented with data that contain $\tau_1 + \tau_2$ periods (equation (3)).

[25] By inspection, we have $g_{irt}(m_t) \leq 0$ for any m_{rt} in that case.

[26] The degenerate solution involves $m_t^* = \tilde{m}_t = 0$.

Under Assumption 1, the leader sets the supermarkups to maximize its profit, and there are region-specific first order conditions:.

$$\tilde{h}_{rt}(m_{rt};\Psi_t) \equiv \frac{\partial \pi^{PL}_{1rt}(m_{rt};\Psi_t)}{\partial m_{rt}} = 0 \quad \text{for } r = 1, 2, \ldots, R \tag{9}$$

By inspection, the supermarkups that solve the leader's problem depend on the economic state, Ψ_t, but not on the timing parameter, η. The timing parameter enters the maximization problem only through the constraints, which do not bind in this case.

Under Assumption 2 there is a binding constraint, and the solution is characterized by the constraint itself and a series of $R-1$ Kuhn-Tucker balancing equations. Without loss of generality, let the firm with the binding IC constraint be firm $k \in \mathbb{C}$ (this could be any coalition firm, including the leader). Then we obtain:

$$g_{kt}(m_t;\eta,\Psi_t) = 0 \tag{10}$$

and, for each $r = 1, 2, \ldots, R-1$,

$$h_{rt}(m_t;\Psi_t) \equiv \frac{\frac{\partial \pi^{PL}_{1rt}(m_{rt};\Psi_t)}{\partial m_{rt}}}{\frac{\partial \pi^{D,k}_{krt}(m_{rt};\Psi_t)}{\partial m_{rt}} - \frac{\partial \pi^{PL}_{krt}(m_{rt};\Psi_t)}{\partial m_{rt}}} - \frac{\frac{\partial \pi^{PL}_{1Rt}(m_{Rt};\Psi_t)}{\partial m_{Rt}}}{\frac{\partial \pi^{D,k}_{kRt}(m_{Rt};\Psi_t)}{\partial m_{Rt}} - \frac{\partial \pi^{PL}_{kRt}(m_{Rt};\Psi_t)}{\partial m_{Rt}}} = 0 \tag{11}$$

Equations (10) and (11) can be obtained with some algebra, after differentiating the Lagrangian of the leader's problem (Appendix B.2). The numerators in the balancing equations are the benefit to the leader of a higher supermarkup; the denominators capture the shadow cost of a higher supermarkup, which arises from the binding firm's IC constraint. The solution depends on both the economic state and the timing parameter.

3.5 Price Leadership Equilibrium

We now formally define the *price leadership equilibrium* (PLE), which is a subgame perfect equilibrium (SPE). The leader's strategy is $\sigma_1 : \mathbb{H} \rightarrow \mathcal{M} \times \mathcal{R}^{J_1}$, where $\mathbb{H}$ is the set of histories, $\mathcal{M}$ is the set of possible supermarkups, and J_1 is the number of products controlled by the leader. The strategies of firms $i = 2, \ldots, F$ are $\sigma_i : \mathcal{M} \times \mathbb{H} \rightarrow \mathcal{R}^{J_i}$.

For simplicity, we first examine Grim Trigger strategies in the infinitely-repeated version of the game, and then map the result into the full model using Proposition 2. In the pricing subgame, let firms price according to p^B_t if $g_{it}(m_t;\eta,\Psi_t) < 0$ for any i or if, in any previous period s, prices differ from $p^{PL}_s(m_s)$. Otherwise let firms price according to $p^{PL}_t(m_t)$.

It is easily verified that no firm can unilaterally improve its payoff in the pricing stage. For example, no firm would set deviation prices, $p_t^{D,i}(m_t)$, because if any firm prefers deviation then this is known by all firms and play shifts to Bertrand. In the announcement stage, let the leader set supermarkups, $m_t^*(\eta, \Psi_t)$, that solve its constrained maximization problem, taking as given rational expectations and the strategies proposed above for the pricing stage. This maximizes the leader's payoff, by construction, so the leader has no incentive to do otherwise. It also ensures that deviation does not occur in the pricing stage. Thus, the stated strategies constitute an SPE, and we label it the PLE.

An implication of Proposition 2 is an alternative set of strategies featuring τ_1 periods of deviation and τ_2 periods of punishment supports identical play along the equilibrium path. A full characterization of this equilibrium is unnecessary for our purposes, though we note that some pricing friction must exist in order to explain why punishment would not ensue immediately after deviation. Finally, it is worth highlighting that the PLE might not be Pareto optimal for the coalition firms because the leader acts in its own interest and side payments—which violate antitrust statutes—are not incorporated.[27]

3.6 Relationship to the Theoretical Literature

The price leadership model resembles the canonical Rotemberg and Saloner (1986) model of collusion because information is perfect and deviation does not occur along the equilibrium path. The main distinction is that we incorporate the idea that price signaling can help support supracompetitive prices. The conditions under which it is reasonable to assume that beliefs respond to cheap talk, such as the leader's price announcement, are discussed in the literature (e.g., Aumann (1990), Farrell and Rabin (1996)).[28] Harrington et al. (2016) develops experimental evidence that price announcements can help facilitate coordination in repeated oligopoly games, and qualitative evidence previously described provides support for our approach in the context of the beer industry.

To incorporate multiple regions and the possibility that multi-market contact affects equilibrium outcomes, we extend the two-firm, two-region model of Bernheim and Whinston (1990).[29] We maintain the assumption that IC constraints are pooled across regions because

[27] See Asker (2010) and Asker et al. (2019) for two empirical examples of inefficient coordination.

[28] In our model, the announcement is "self-committing" because the leader has no incentive to deviate from a perfect equilibrium. It is not "self-signaling" because the leader would prefer the followers to accept the supermarkup even if it plans to deviate. Farrell and Rabin (1996) state that "a message that is both self-signaling and self-committing seems highly credible" yet point to an experimental literature to support that cheap talk can be effective in shaping beliefs even if not self-signaling.

[29] See Section 7 on differentiated products. The necessary conditions are analogous to the Kuhn-Tucker

it generates leader profit at least weakly larger than the alternative of independent regions. That said, the feasibility of the pooled solution depends on the sophistication of firms, into which we have little visibility (insofar as it relates to multi-market contact). It is also possible to make empirical progress with independent regions, as the supermarkups that solve the constrained maximization problem would be determined by

$$g_{krt}(m^*_{rt}; \eta, \Psi_t) = 0 \tag{12}$$

where firm $k \in \mathbb{C}$ has the binding constraint. Later in this paper, we explore the implications of multi-market contact in a counterfactual simulation. We also provide imputation results for the case of region-specific IC constraints (Appendix D).

The price leadership model incorporates firm heterogeneity, which is important for applied work in most industries. This raises questions about which firms participate in the coalition and, among these, which is the leader. We impose *ex ante* that the coalition includes ABI, Miller, and Coors (or MillerCoors) and that ABI is the leader, an approach that is supported by the available qualitative and empirical evidence. In principle, a more theoretical approach to the coalition could be taken, under the assumption that each firm faces a decision whether to join the coalition (e.g., as in d'Aspremont et al. (1983), Donsimoni et al. (1986), and Bos and Harrington (2010)). Similarly, in price leadership models slightly different than ours, Pastine and Pastine (2004) allows a war of attrition to determine the leader, and Ishibashi (2008) and Mouraviev and Rey (2011) show that coalition profits are higher if the leader is the firm with the greatest incentive to deviate.[30] As multiple theoretical assumptions appear to be available, we prefer the empirical approach.

Finally, our model abstracts away the retail and distribution sectors. However, it is isomorphic to a model that incorporates constant markups, or "cost-plus" pricing, downstream. The reason is that downstream markups and brewer marginal costs enter the profit functions in the same way, so that downstream markups are equivalent to a tax on production (Appendix F). Recent research provides some support for cost-plus pricing among retailers in scanner data similar to ours.[31] The model would be misspecified for settings in which retailers exercise buyer power to obtain lower prices (e.g., as in Loertscher and Marx

balancing equations that characterize the solution to the leader's constrained maximization problem.

[30] Our model differs in that each period features an announcement followed by simultaneous pricing, rather than sequential pricing.

[31] DellaVigna and Gentzkow (2019) shows that retail prices often do not respond to local demand shocks, and Butters et al. (2020) documents that retail prices change one-for-one with local cost shocks (generated by excise taxes). This combination would arise from cost-plus pricing.

(2019)). In our setting, buyer power may be limited by the lack of private label store brands and the regulatory prohibition on slotting allowances, which makes it harder for retailers to discipline coordination by auctioning shelf space. Nonetheless, downstream markups appear to be sizable. Tremblay and Tremblay (2005, pp. 161-162) place them at about 35% of total revenue, based on industry studies conducted in the 1970s and 1990s, such that brewer revenues would be about 65% of total revenue.

4 Empirical Implementation

4.1 Assumptions and Imputation Algorithms

We make a number of assumptions to map the data into the model of price leadership. First, we assume that firms set prices according to the PLE, so that the price vector $p_t^{PL}(\Psi_t)$ is observed in the data. Second, we assume that the coalition comprises ABI (the leader) and either Miller and Coors (in 2006 and 2007) or MillerCoors (in 2010 and 2011). These first two assumptions are supported by the qualitative evidence in Section 2. Third, we assume that demand is given by the RCNL-2 model estimated in Miller and Weinberg (2017). Fourth, we assume that each period ($t = 1, \ldots, \infty$) in the theoretical model corresponds to a fiscal year in the data. Our treatment incorporates that demand and cost conditions change within the fiscal year, while the supermarkups are fixed. As the empirical implementation uses the equivalent slack function, this also implies that deviation lasts for one year before punishment ensues. If deviation is longer or shorter than one year in reality, this would be subsumed into the timing parameter.

Marginal costs, Bertrand prices, and the supermarkups can be recovered from the data by applying the structure of the model. We have already described how marginal costs and Bertrand prices can be obtained given the supermarkup (Proposition 1). Thus, we focus on inferring the supermarkups from the leader's constrained maximization problem. Recall that the leader's first order conditions depend on (i) whether an IC constraint binds and (ii) in the case of a binding constraint, the value of the timing parameter. Our general approach is to make *ex ante* assumptions on (i) and (ii) and then evaluate these assumptions *ex post*. The algorithms that recover the supermarkup depend on the assumptions.

The constrained case is the most demanding from a computational standpoint because it requires solving a system of R nonlinear equations with R unknowns (with $R = 37$ in our application). We initially consider $\eta = (0.20, 0.25, 0.30, 0.35, 0.40)$, based on our experience with the data and the model. For each of these timing parameters in turn, we apply the following algorithm to each period:

1. Consider a candidate supermarkup vector, $\widehat{m}_t$.

2. Obtain implied marginal costs and Bertrand prices (Proposition 1).

3. Obtain $g_{it}(\widehat{m}_t; \eta, \Psi_t)$ for all $i \in \mathbb{C}$ by computing deviation prices using the static first order conditions (equation (2)), calculating profit under price leadership, deviation, and Bertrand, and then applying the proposed timing parameter.

4. Identify the binding firm, k, such that $g_{kt}(\widehat{m}_t; \eta, \Psi_t) = \min_{i \in \mathbb{C}} g_{it}(\widehat{m}_t; \eta, \Psi_t)$.

5. Obtain $h(\widehat{m}_{rt}; \Psi_t)$ for $r = 1, \ldots, R-1$, by numerically differentiating the price leadership profit of the leader and the binding firm, as well as the deviation profit of the binding firm, with respect to the supermarkups.

6. Assess the loss function, which we construct based on the observation that if $\widehat{m}_t = m_t$ then $g_{kt}(\widehat{m}_t; \eta, \Psi_t) = 0$ and $h_{rt}(\widehat{m}_{rt}; \Psi_t) = 0$ for $r = 0, 1, \ldots, R-1$.

7. Update $\widehat{m}_t$ if needed, and repeat to convergence.

The computational burden of such an algorithm typically increases nonlinearly in the number of unknowns (here, supermarkups). Thus, we develop an approach that exploits the structure of the model, and which makes the computational burden roughly linear in the number of supermarkups. The details are provided in Appendix C.

The unconstrained case is simpler computationally because each region can be considered in isolation, yielding R distinct problems, each with one equation and one unknown. We apply a standard equation solver to identify the supermarkup, yielding $\tilde{h}_{rt}(m_{rt}; \Psi_t) = 0$. Finally, we also consider the case of Bertrand competition, for which marginal costs can be recovered from the static first order conditions of equation (2).

4.2 Identification

Each of the *ex ante* modeling assumptions considered above implies corresponding supermarkups and marginal costs, and this provides a path to *ex post* model selection. We observe that timing parameters closer to one imply higher supermarkups because they create more slack in the IC constraints. In turn, higher supermarkups imply lower marginal costs for coalition firms, holding fixed the observed prices. We illustrate this latter point in Figure

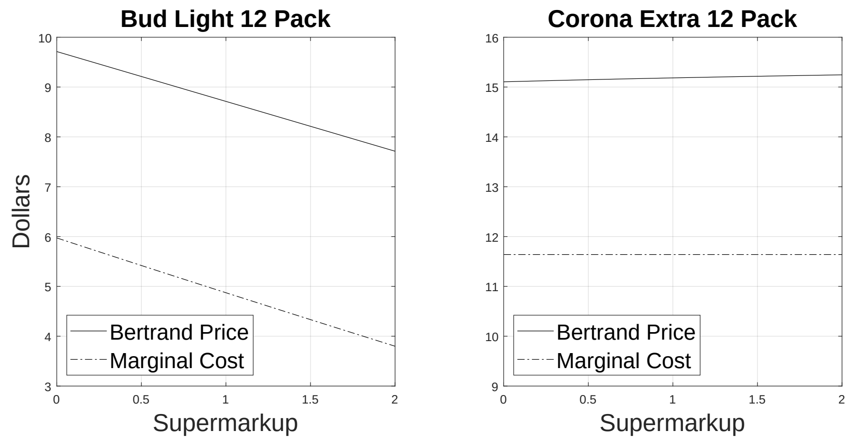

Figure 2: Bertrand Prices and Marginal Costs Under Different Supermarkups

Notes: The figures plot the average Bertrand prices and supermarkups that one would infer given candidate supermarkups in the range of $m \in [0, 2]$. The statistics are obtained for Bud Light 12 packs (left panel) and Corona Extra 12 packs (right panel) in the fiscal year 2007.

2, which plots the Bertrand prices and marginal costs one would infer from different supermarkups. Putting these observations together, in our application there is a one-to-one mapping between the timing parameter and both (i) the costs of coalition firms and (ii) the relative costs of coalition and fringe firms. This can allow the relative quality of the *ex ante* modeling assumptions to be evaluated.[32]

In particular, if one has prior knowledge of marginal costs, then the timing parameter and the associated supermarkups are identified. We see at least two workable approaches. The first exploits engineering or business data on the magnitude of marginal costs (e.g., as in Nevo (2001)), and simply selects the *ex ante* modeling assumptions that imply the correct marginal costs. This does not require multiple periods or regions, and may be particularly useful in merger review because antitrust authorities can use subpoena power to gain access to confidential business documents. Unfortunately, we do not have access to publicly-available information on the marginal costs of the beers in our sample.

We take a second approach, which involves applying prior knowledge about how marginal costs change with the economic state. To implement, we specify the following marginal cost

[32]We suspect this identification argument would extend in similar applications. Fringe costs are invariant to the candidate supermarkup, which can be seen in Figure 2 and verified with the proof to Proposition 1.

function for product j in region r, fiscal year t, and quarter s:

$$\begin{aligned} mc_{jrts} &= \beta_1 \mathbb{1}\{j \in \text{ABI}\} \times \mathbb{1}\{t \in (2010, 2011)\} \\ &+ \beta_2 \mathbb{1}\{j \in \text{Miller}\} \times \mathbb{1}\{t \in (2010, 2011)\} \\ &+ \beta_3 \mathbb{1}\{j \in \text{Coors}\} \times \mathbb{1}\{t \in (2010, 2011)\} \\ &+ \beta_4 \text{ Distance}_{jrts} + \mu_j + \mu_r + \mu_{ts} + \epsilon_{jrts} \end{aligned}$$

The first three terms allow for the marginal costs of ABI, Miller, and Coors products to shift after the Miller/Coors merger. Distance is between the brewery and the region in question, and accounts for transportation costs. The terms μ_j, μ_r, and μ_t are product, region, and time (year$\times$quarter) fixed effects, respectively. Of particular relevance for our identification strategy, the time effects account for any changes in the prices of labor and ingredients that affect all firms equally.[33]

Our identifying assumption is $\beta_1 = 0$, which yields the marginal cost function of Miller and Weinberg (2017). Given the presence of product and time fixed effects, the identifying assumption implies that ABI's marginal costs do not change differently from those of Modelo and Heineken, on average, with the 2008 Miller/Coors merger. Equivalently, recalling Figure 1, the assumption is that the true model generates the ABI price increases that are observed in the data after the Miller/Coors merger without ABI-specific cost increases.

We estimate the marginal cost function with OLS after stacking the observations from each fiscal year. This allows us to evaluate the *ex ante* modeling assumptions on the IC constraints, and reject those that generate a $\hat{\beta}_1$ that is statistically different than zero. Further, we select as the preferred assumptions those that generate the $\hat{\beta}_1$ that is closest to zero. In principle, one could estimate the timing parameter using a nested method-of-moments procedure, but that is computationally infeasible in our context.[34]

We view the identifying assumption as a reasonable approximation given the institutional details of the market. There are a number of reasons the assumption might not hold exactly. Changes in the prices of ingredients (e.g., water, hops, barley, and yeast) would affect ABI costs differently than import costs if ingredients are used in different proportions. Similarly, it is feasible that the wages paid by ABI could trend differently from those paid by Modelo and Heineken, especially as their brewing occurs in different countries. However,

[33]There are four fiscal years, each with four quarters, so there are 16 time fixed effects.

[34]If we run the imputation algorithm for constrained price leadership on a single processor, it takes three to four weeks converge, depending on the timing parameter. A number of factors contribute to the time costs, including the number of regions and the 500 demographic draws used in the RCNL demand system. We reduce the computational burden somewhat by using different processors for different fiscal years.

we are skeptical that input prices changes would substantially impact the validity of the identifying assumption. For example, with respect to ingredients, the brands in our sample are all simple lagers, so proportions are likely to be roughly similar and, furthermore, some information suggests that ingredient and labor costs are small relative to the retail price.[35]

Also worth discussing at this point is the InBev acquisition of Anheuser-Busch, which closed in 2009. As best we can discern, the post-acquisition efforts to reduce costs would not have affected marginal costs. InBev revised the pay system, ended pension contributions and life insurance for retirees, and transferred all foreign beer operations to InBev (Ascher (2012)). The domestic distribution of the brands in our sample was unaffected, so transportation cost would not have changed, and in any event the cost specification controls for distance. An interesting possibility is that the acquisition could have affected buyer power in the hops market, reducing ABI's input costs relative those of its competitors.[36] However, corroborating our understanding of the acquisition, we have been unable to find references to substantial variable costs changes in the ABI annual reports or the popular press.

5 Results

5.1 Model Selection

Table 3 summarizes the results of the model selection exercise. Each column corresponds to one set of *ex ante* modeling assumptions. Results for Bertrand and unconstrained price leadership are shown in the left-most and right-most columns, respectively, and results for constrained price leadership under various timing parameters are shown in the middle columns. Panel A provides the main regression coefficients that we obtain from OLS estimation of the marginal cost function. Standard errors are clustered at the region level and shown in parentheses. Panel B provides the average supermarkups, $\bar{m}_t = (1/R)\sum_r m_{rt}$, and the proportion of marginals costs that are negative.

The timing parameter of $\eta = 0.26$ generates the $\hat{\beta}_1$ closest to zero, and so it is our preferred model.[37] The Bertrand and unconstrained price leadership models are easily re-

[35]We draw on Tremblay and Tremblay (2005, pp. 161-162) which places ingredients costs at 4-6% of retail revenues and labor costs as 5-12% of retail revenues, based on industry studies conducted in the 1970s and 1990s. We do not know which portion of labor costs should be treated as a marginal cost. According to Tremblay and Tremblay, distribution and packaging accounts for the bulk of brewer costs, and we think the marginal cost specification accounts for those expenditures reasonably well.

[36]Our marginal cost specification accounts for the analogous possibility that the Miller/Coors merger amplified buyer power in the hops market, through the β_2 and β_3 parameters.

[37]After our examination of the initial timing parameters $\eta = (0.20, 0.25, 0.30, 0.35, 0.40)$, we determined

Table 3: Model Selection and Implied Supermarkups

Panel A: OLS Regression Results

	Bertrand	$\eta = 0.20$	$\eta = 0.25$	$\eta = 0.26$	$\eta = 0.30$	$\eta = 0.35$	$\eta = 0.40$	Unconstrained
β_1	0.657	0.204	0.024	-0.017	-0.203	-0.492	-0.858	0.607
	(0.094)	(0.117)	(0.122)	(0.123)	(0.125)	(0.127)	(0.126)	(0.147)
β_2	0.189	-0.286	-0.476	-0.519	-0.716	-1.023	-1.411	-0.017
	(0.074)	(0.096)	(0.102)	(0.103)	(0.107)	(0.113)	(0.137)	(0.137)
β_3	-0.067	-0.557	-0.748	-0.791	-0.987	-1.288	-1.663	-0.242
	(0.095)	(0.107)	(0.113)	(0.114)	(0.120)	(0.129)	(0.173)	(0.173)
β_4	0.220	0.229	0.234	0.235	0.240	0.251	0.269	0.340
	(0.066)	(0.065)	(0.066)	(0.067)	(0.069)	(0.073)	(0.082)	(0.120)

Panel B: Other Statistics

	Bertrand	$\eta = 0.20$	$\eta = 0.25$	$\eta = 0.26$	$\eta = 0.30$	$\eta = 0.35$	$\eta = 0.40$	Unconstrained
$\overline{m}_{2006}$	0	0.84	1.10	1.15	1.39	1.71	2.07	4.51
$\overline{m}_{2007}$	0	0.87	1.14	1.20	1.44	1.77	2.15	4.65
$\overline{m}_{2010}$	0	1.27	1.70	1.80	2.22	2.83	3.57	4.64
$\overline{m}_{2011}$	0	1.31	1.75	1.85	2.28	2.90	3.64	4.79
$mc < 0$	<0.001	0.002	0.005	0.006	0.011	0.023	0.050	0.170

Notes: Panel A summarizes the results from OLS estimation of the marginal cost function, based on the marginal costs obtained under different modeling assumptions. There are 20,162 observations at the product-region-fiscal year-quarter level, combining data from fiscal years 2006, 2007, 2010, and 2011. The intermediate timing parameters are generated under the assumption that the IC constraint binds. Regressors include indicators for ABI brands, Miller brands, and Coors brands, in the fiscal years 2010 and 2011 (corresponding to β_1, β_2, and β_3), distance from the brewery to the region (β_4), as well as product, region, and time fixed effects. Standard errors are clustered at the region level and shown in parentheses. Panel B provides the average supermarkups by fiscal year, and the proportion of marginal cost values which are negative.

jected. There are timing parameters near $\eta = 0.26$ for which the null of $\beta_1 = 0$ cannot be rejected, with $\eta = 0.20$ and $\eta = 0.30$ being near the boundaries of this range.[38] Among the constrained models, the average supermarkups increase with the timing parameter, as does the average change before versus after the Miller/Coors merger. This pattern arises because more slack is inferred in IC constraints (for any given price) if the timing parameter is greater. At the preferred model, we infer average supermarkups of \$1.15, \$1.20, \$1.80, and \$1.85 in the four fiscal years, respectively.[39]

Turning to the remaining parameters, we estimate that the marginal cost intercepts of

---

that $\eta = 0.26$ would probably generate the OLS estimate of β_1 closest to zero, and indeed that is the case.

[38] The relevant t-statistics are -1.74 for the $\eta = 0.20$ model and 1.62 for the $\eta = 0.30$ model.

[39] Unconstrained price leadership generates the largest implied supermarkups. These supermarkups do not change much with the Miller/Coors merger because there is only a small effect on ABI's preferred supermarkup. This also explains why $\hat{\beta}_1 > 0$ for that model—without an increase in supermarkups, the model requires ABI-specific marginal cost increases to rationalize the price patterns of Figure 1.

Table 4: Welfare Effects of Price Leadership

	2006	2007	2010	2011
Producer Surplus				
Δ Total Profit	17.28%	17.43%	22.18%	22.10%
Consumer Surplus (CS)				
ΔCS/ΔTotal Profit	-1.56	-1.54	-1.70	-1.72

Notes: The table summarizes the effects of price leadership, based on a comparison of the observed equilibrium to a counterfactual that features Bertrand prices. The profit statistics are calculated by aggregating product-region-quarter observations, subtracting Bertrand profit from price leadership profit and then dividing by Bertrand profit. The consumer surplus statistics are calculated by aggregating over region-quarter combinations, and then normalizing by the change in total profit.

Miller and Coors decrease with the joint venture by \$0.52 and \$0.79 (β_2 and β_3) with the preferred model. As the distance estimate is positive (β_4), a second source of efficiencies from Miller/Coors arises as production of Coors brands and, to a lesser extent Miller brands, is moved to breweries closer to retail locations. Miller and Weinberg (2017) estimates similar marginal cost parameters and analyzes the efficiencies in greater detail.

Table 4 analyzes the welfare consequences of price leadership. We benchmark against Bertrand equilibrium, which we compute holding fixed the imputed marginal costs from the preferred model. Across all firms, we find that the difference in profit under price leadership and Betrand is about 17% of Bertrand profits in fiscal years 2006 and 2007, and 22% in fiscal years 2010 and 2011. The adverse impact of price leadership on consumer surplus is about 1.5 times larger than this profit gain in 2006 and 2007, and about 1.7 times larger in 2010 and 2011.[40] Thus, the loss of total welfare is just greater than half of the profit gain.

Table 5 provides the average markup for each product in the data in fiscal years 2007 and 2010, again based on the preferred model. Across all 20,162 brand-size-region-quarter-year observations, the median markup is \$4.53 on an equivalent-unit basis, and accounts for 44% of the retail price.[41] The average markups on ABI 12 packs tend to be about \$0.55 higher in 2010 as compared to 2007, which reflects higher Bertrand prices and supermarkups. The markups on Miller 12 packs tend to be about \$1.20 higher and the markups on Coors

[40]We report effects in this manner because consumer surplus is identified only up to an additive constant, so we cannot recover percentage changes.

[41]Tremblay and Tremblay (2005, pp 161-162) report that downstream markups account for about 35% of total revenue. The brewer markups we recover would account for 63% of the remainder, on average, which would correspond to the margins or Lerner Index. In our experience, this is somewhat high but not exceptional for consumer products, and is certainly plausible with supracompetitive pricing. As a point of comparison, Nevo (2001) reports margin estimates of 64.4% and 57.4% for manufacturers in the ready-to-eat cereals industry, based on the Annual Survey of Manufacturers and an industry study, respectively.

Table 5: Brewer Markups

Brand	6 Packs 2007	6 Packs 2010	12 Packs 2007	12 Packs 2010	24 Packs 2007	24 Packs 2010
Bud Light	5.19	5.73	5.04	5.61	4.92	5.48
Budweiser	5.35	5.87	5.19	5.74	5.04	5.65
Coors	4.15	5.72	3.97	5.61	3.82	5.45
Coors Light	3.90	5.55	3.77	5.44	3.66	5.33
Corona Extra	3.65	3.28	3.36	2.99	3.34	3.06
Corona Light	3.41	3.09	3.09	2.77	3.17	2.91
Heineken	3.53	3.28	3.28	3.00	3.34	3.26
Heineken Light	3.16	2.97	2.86	2.66	2.97	2.90
Michelob	5.27	5.94	5.19	5.84	4.39	5.70
Michelob Light	5.19	5.75	5.06	5.62	5.00	5.24
Miller Gen. Draft	4.42	5.61	4.24	5.47	4.13	5.36
Miller High Life	4.43	5.65	4.24	5.46	4.15	5.38
Miller Lite	4.39	5.58	4.25	5.48	4.13	5.35

Notes: This table provides the average markups for each product (brand–size combination) in fiscal years 2007 and 2010.

products tend to be about \$1.67 higher, reflecting the combined impact of higher Bertrand prices, higher supermarkups, and lower marginal costs.

Before proceeding, it is worth discussing whether the timing parameter that emerges from our analysis, $\eta = 0.26$, comports with reasonable priors about oligopoly supergames. As a purely mathematical observation, it is not difficult to generate such a timing parameter if the duration of deviation exceeds that of punishment. For example, the combination $(\delta = 0.9, \phi = 0.9, \tau_1 = 2, \tau_2 = 1)$ yields $\eta = 0.27$, through an application of equation (7). We are agnostic about whether deviation might indeed last longer than punishment, especially as firms seemingly should not punish longer than is necessary, and so view our result as plausible. Higher timing parameters obtain if punishment is relatively longer than in the above example. The combination $(\delta = 0.9, \phi = 0.9, \tau_1 = 1, \tau_2 = 1)$ yields $\eta = 0.45$ and the combination $(\delta = 0.9, \phi = 0.9, \tau_1 = 1, \tau_2 = 5)$ yields $\eta = 0.74$.

An alternative interpretation of our timing parameter is that it embeds the influence of some pricing friction that is not modeled explicitly. Different micro-foundations are available—the friction could arise if the leader is not certain about the IC constraints of other firms and wishes to avoid an accidental punishment phase, or if price leadership creates some risk of antitrust penalties, for example. Let the augmented slack function be

$$g_i \equiv \frac{1}{1-\eta^*}\pi_i^{PL} - \pi_i^{D,i} - \frac{\eta^*}{1-\eta^*}\pi_i^B - \zeta \qquad (13)$$

where ζ is the friction. There is a continuum of (η, ζ) combinations that would generate a binding IC constraint of $g_i = 0$. To illustrate, we express the timing parameter in terms of the friction:

$$\eta^* = \frac{\pi_i^{D,i} - \pi_i^{PL} + \zeta}{\pi_i^{D,i} - \pi_i^{B} + \zeta} \tag{14}$$

Our baseline analysis implicitly imposes $\zeta = 0$ and obtains $\eta = 0.26$. However, these two objects are not separately identifiable, and the presence of a friction would lead us to understate the timing parameter. This would not affect our imputation results or the analysis of equilibrium because the binding IC constraint crosses zero at the same supermarkup—indeed that is what defines the continuum of jointly identified (η, ζ) combinations.[42] It could generate different counterfactual inferences if the friction depends on the supermarkup.

5.2 Analysis of Equilibrium

We now examine the preferred model in greater detail. To start, we provide some intuition for how the supermarkups are selected and why they change with the Miller/Coors merger. We exploit that it is possible to compute profit under price leadership, deviation, and Bertrand $(\pi^{PL}, \pi^{D}, \pi^{B})$ for any counterfactual supermarkups, holding fixed marginal costs that are recovered with the preferred model. We consider counterfactual supermarkups $\tilde{m}_t = \iota m_t$, for $\iota = (0, 0.01, 0.02, \ldots, 1.50)$. This scales all of the region-specific supermarkups by the same multiplicative factor; with $\iota = 0$ the outcomes are identical to Bertrand, and with $\iota = 1$ the supermarkups are those that arise in the PLE.

Figure 3 plots price leadership and deviation profit as indices relative to Bertrand profit. Results are provided for fiscal years 2007 and 2010, which immediately predate and postdate the Miller/Coors merger. The vertical blue line marks the PLE (equivalently, $\iota = 1.00$). The profit functions take a value of one at $\tilde{m}_t = 0$ because outcomes are equivalent to Bertrand. From there, they increase in the supermarkup, with deviation profit increasing at a faster rate than price leadership profit. If we were to extend these graphs to large enough supermarkups, profit under price leadership would flatten and (eventually) start to fall, whereas the slope of deviation profit would converge to zero.

At the PLE in fiscal year 2007, price leadership increases profit above Bertrand by 15% for ABI, 20% for Miller, and 25% for Coors. Deviation increases static profit even more, and this is especially true for Coors, for which deviation profit exceeds Bertrand profit by 35% at

[42] In Appendix Figure G.4, we plot the (η, ζ) combinations under which the MillerCoors IC constraint binds in fiscal year 2010. With $\eta = 0.45$, the magnitude of the friction is equivalent to 8.25% of MillerCoors' profit in the PLE.

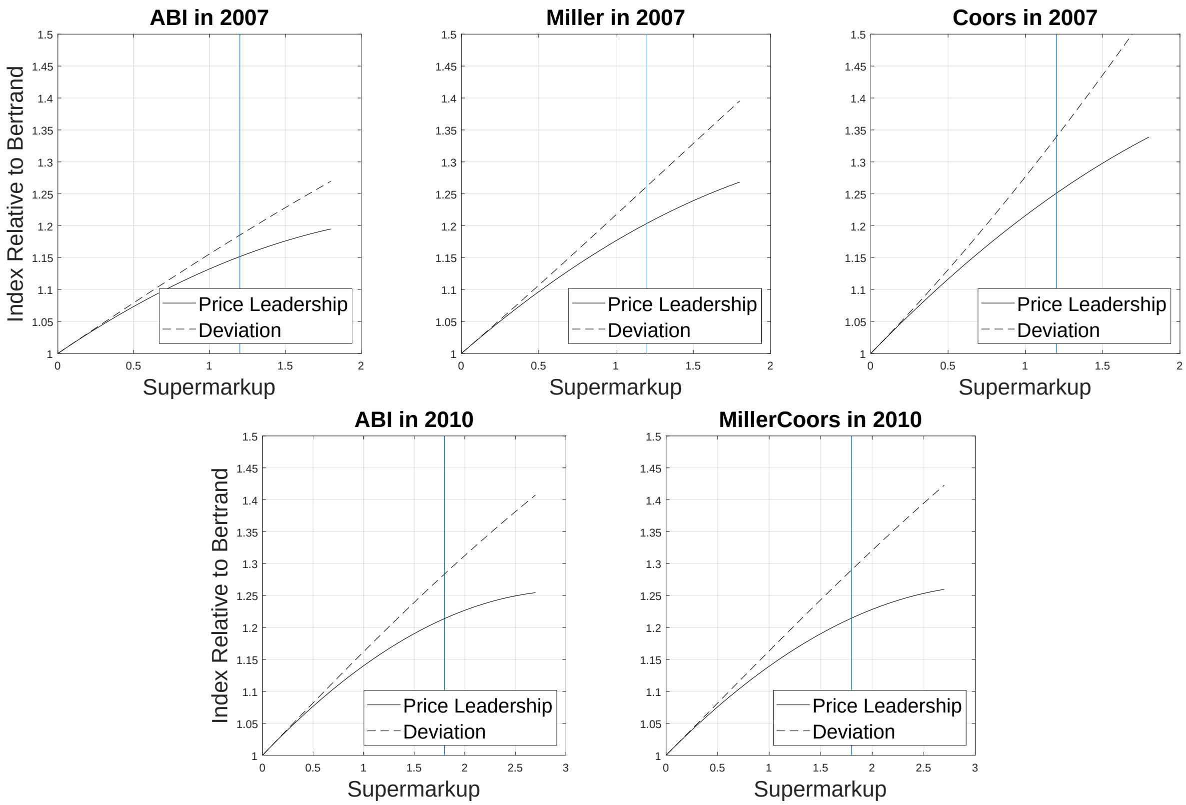

Figure 3: Profit Functions of Coalition Firms

Notes: The figure provides the profit under price leadership and deviation for coalition firms in fiscal years 2007 and 2010, reported as an index relative to Bertrand profit. Statistics are computed for a range of supermarkups. At the vertical blue line, the region-specific supermarkups are those recovered from the imputation procedure with a timing parameter of 0.26 ($\overline{m}_{2007} = 1.20$ and $\overline{m}_{2010} = 1.80$). These supermarkups are scaled by the multiplicative factors $\iota = (0.00, 0.01, \ldots, 1.50)$ in the evaluation of the profit functions.

the PLE. Thus, among the coalition firms, it appears that Coors has the greatest incentive to deviate, and we verify this momentarily. There is more symmetry in fiscal year 2010. Price leadership increases the profit of ABI and MillerCoors by 21%, relative to Bertrand, evaluated at the PLE, and deviation increases profit by 29% for both firms.

Figure 4 plots the equivalent slack functions for the coalition firms in fiscal years 2007 and 2010, which we obtain from equations (5) and (6) using the timing parameter of 0.26 and the profit functions just analyzed. The vertical blue line marks the PLE. The IC constraints in the leader's maximization problem are satisfied if all of the slack functions are weakly positive (i.e., at or above the horizontal blue line). The slack functions are positive for small supermarkups and negative for large supermarkups. In 2007, the slack function of Coors crosses zero at the PLE, even as the slack functions of Miller and ABI are positive. Thus,

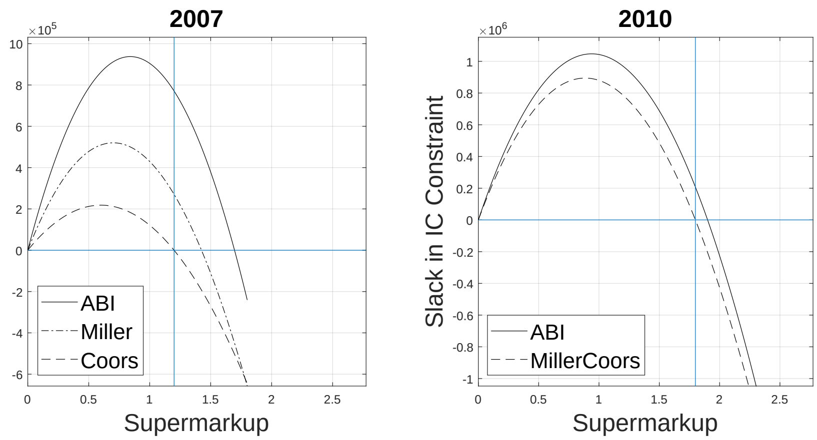

Figure 4: Slack Functions of the Coalition Firms

Notes: The figure provides the equivalent slack functions of the coalition firms in each fiscal year. At the vertical blue lines, the region-specific supermarkups are those recovered from the imputation procedure with a timing parameter of 0.26 ($\overline{m}_{2007} = 1.20$ and $\overline{m}_{2010} = 1.80$). These supermarkups are scaled by the multiplicative factors $\iota = (0.00, 0.01, \ldots, 1.50)$ in the evaluation of the slack functions.

Coors provides the binding constraint. The MillerCoors slack function (in 2010) crosses zero with larger supermarkups than either the Miller or Coors slack functions (in 2007), which explains why average supermarkups are higher in 2010 than in 2007. This suggests that the Miller/Coors merger is the main driver of the higher supermarkups in the latter half of our sample, though the illustration is not definitive because demand and cost conditions also change. We use a counterfactual later in the paper to isolate the effect of the merger.

We turn now to the dispersion of supermarkups across regions. Figure 5 shows that supermarkups tend to be higher in regions where ABI has a large market share and Coors (in 2007) or MillerCoors (in 2010) have a smaller market share. This reflects the Kuhn-Tucker conditions that characterize the solution to ABI's constrained maximization problem: ABI benefits more from a higher supermarkup if it has a large market share, and the effect of a higher supermarkup on the binding IC constraint is greater if Coors and MillerCoors have large market shares. As the HHI tends to be high if ABI is large and other firms are small, the region-specific supermarkups are also correlated with the HHI. These are not structural relationships, but are informative nonetheless because market shares provide some composite measure of consumer willingness-to-pay and marginal cost.

To explore the role of multi-market contact in generating this variation across regions, we recompute equilibrium under the alternative assumption that the leader faces multiple, distinct, region-specific constrained maximization problems. This amounts to finding the

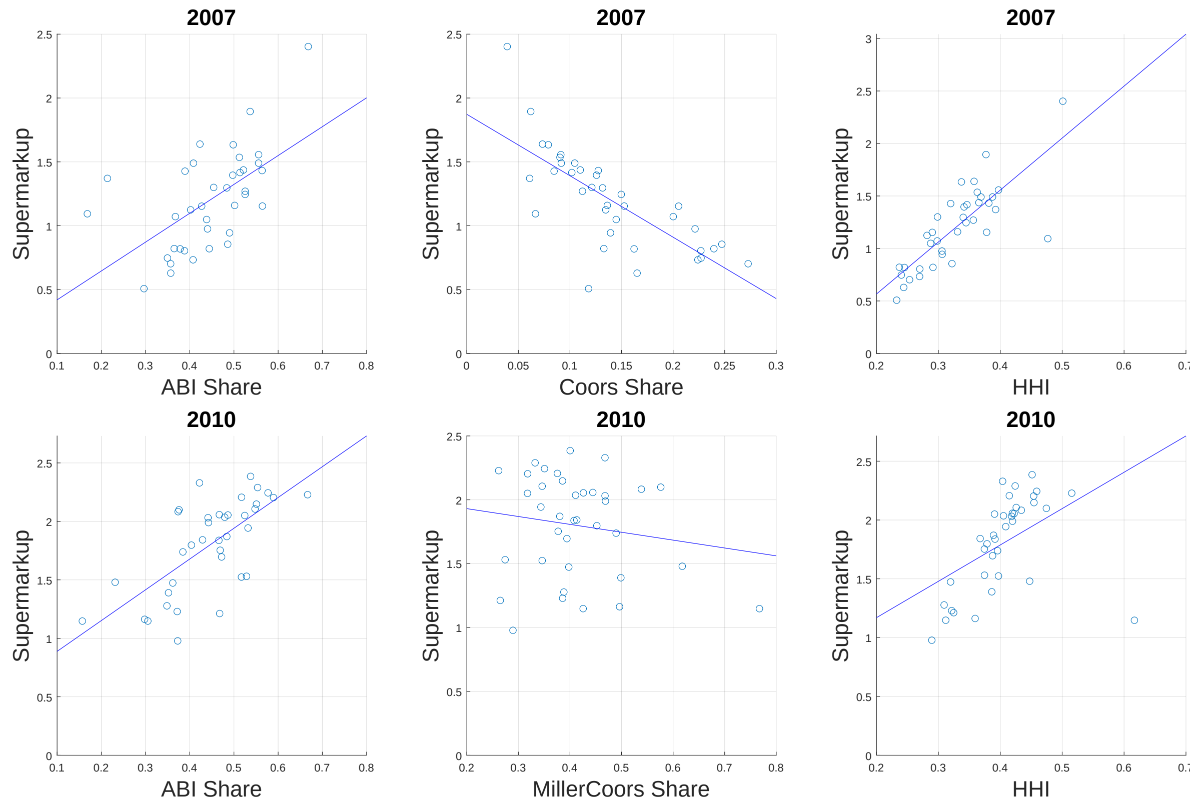

Figure 5: Region-Specific Supermarkups in 2007 and 2010

Notes: The figure provides scatter plots of the region-specific supermarkups in fiscal years 2007 and 2010 against the ABI market share (in 2007 and 2010), the Coors market share (in 2007), the MillerCoors market share (in 2010), and the HHI (in 2007 and 2010). Lines of best fit are provided to assist with interpretation.

supermarkup, m_r, for each region r, that satisfies $g_{irt}(m_{rt}; \eta, \Psi_t) = 0$. We hold fixed the marginal costs and timing parameter that we recover from the baseline model. Figure 6 provides two scatter plots of the counterfactual supermarkups (vertical axis) and the baseline supermarkups (horizontal axis). There are noticeable differences in fiscal year 2007, but these mostly disappear in fiscal year 2010, reflecting the greater symmetry among the coalition firms after the Miller/Coors merger. Interestingly, multi-market contact does not affect the average supermarkup or the profitability of price leadership much.[43]

[43] In fiscal year 2007, the profits of ABI, Miller, and Coors, are about 0.5%, 0.5%, and 1.0% higher, respectively, in the baseline model than in the counterfactual with independent regions. In fiscal year 2010, the profit of both ABI and MillerCoors is about 0.5% higher in the baseline model. In Appendix D, we explore imputation under the alternative assumption that IC constraints are not pooled across regions.

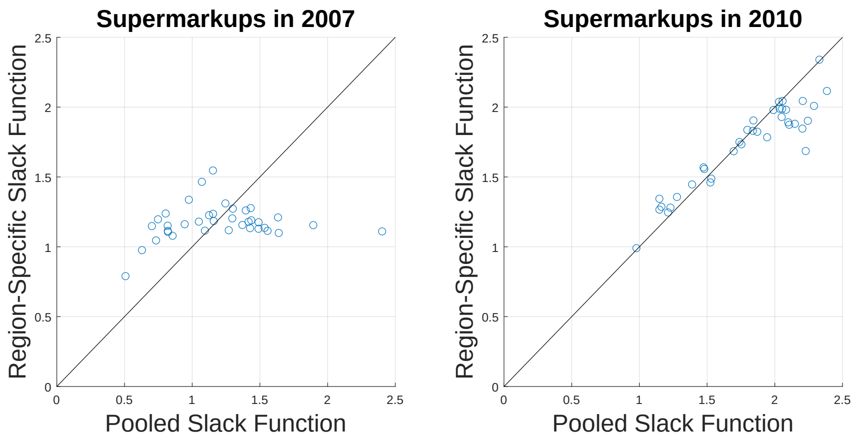

Figure 6: Region-Specific Supermarkups and Multi-Market Contact

Notes: The figure provides scatter plots of the region-specific supermarkets in fiscal years 2007 and 2010 for the baseline model (horizontal axis) and a counterfactual simulation in which the leader solves region-specific constrained maximization problems. Each panel has a 45° line to assist with interpretation.

6 Application to Merger Review

6.1 Background

In this section, we analyze the Miller/Coors and ABI/Modelo mergers using the price leadership model. This covers two interesting merger scenarios. First, with Miller/Coors, the merger involves two coalition firms, including the binding firm. This combination creates slack in the IC constraints at pre-merger prices, and allows the leader to support higher supermarkups in equilibrium. Second, with ABI/Modelo, the merger involves the acquisition of an important fringe firm by the leader. By raising the prices of the acquired fringe firm, the leader can create slack in the binding IC constraint, and thereby support higher supermarkups in equilibrium.[44] We adopt standard antitrust parlance and refer to the change in the supermarkup as the coordinated effect of the merger. The overall price changes also reflect shifts in underlying Bertrand prices, which we refer to as the unilateral effect of the merger. We consider the mergers both with and without efficiencies, in the form of marginal cost reductions.

[44]This mechanism is consistent with the allegations of the DOJ regarding ABI/Modelo:

> ABI and MillerCoors often find it more profitable to follow each other's prices than to compete aggressively.... In contrast, Modelo has resisted ABI-led price hikes.... If ABI were to acquire the remainder of Modelo, this competitive constraint on ABI's and MillerCoors' ability to raise their prices would be eliminated.

See Paras 3-5 of the Complaint in *US v. Anheuser-Busch InBev SA/NV and Grupo Modelo S.A.B. de C.V.*

6.2 Implementation

We focus on fiscal year 2010 for Miller/Coors, which is the first complete fiscal year after the merger. The data contain PLE outcomes with merger efficiencies. We obtain the "no merger" scenario by simulating the PLE under the assumptions that (i) the Miller and Coors brands are owned by separate firms, and (ii) the estimated changes to the Miller and Coors marginal cost functions do not occur. The latter of these involves adding $\hat{\beta}_2$ and $\hat{\beta}_3$ to the marginal costs of Miller and Coors products (see Table 3), and accounting for the greater shipping distances that arise with separate ownership. We also obtain a "merger without efficiencies" scenario by simulating the PLE with joint pricing but no merger efficiencies. A comparison of these scenarios isolates the impact of the merger.

For ABI/Modelo, we focus on fiscal year 2011, which is the closest year in our sample to the acquisition date. The data contain the "no merger" scenario. We simulate PLE outcomes in which ABI and Modelo brands are owned by the same firm. We consider three specific cases: (i) no merger efficiencies, (ii) a "minor" efficiency comprising a \$0.50 reduction in Modelo costs, and (iii) a "major" efficiency comprising product-specific cost reductions which leave Bertrand prices exactly unchanged due to the merger.[45] A comparison of these scenarios against the baseline modeling results isolates the impact of the merger.

There is an additional question about how to model the post-merger prices of Modelo. Because Modelo sells imported beers that are relatively more expensive than those of ABI and MillerCoors, successful post-merger coordination probably would not require that the same supermarkup apply to Modelo brands. Instead, we think it more likely that ABI would set Modelo prices to maximize the present value of its profit, accounting for impacts on IC constraints. As computing such an equilibrium would be computationally prohibitive, we work with an approximation in which post-merger Modelo prices are equal to Bertrand prices plus some region-specific amount that applies to all Modelo products. This essentially adds a second set of supermarkups to the model.

Thus, in the post-merger equilibrium, there are $2R$ supermarkups that must be computed. For notational purposes, we denote the supermarkups that apply to ABI and MillerCoors brands as $m^1 = (m^1_1, m^1_2, \ldots, m^1_R)$ and the supermarkups that apply to Modelo products as $m^2 = (m^2_1, m^2_2, \ldots, m^2_R)$. Letting ABI be firm 1 and MillerCoors be firm k, post-

[45]The major efficiency is a multi-product version of the compensating marginal cost reductions derived in Werden (1996). On average, we reduce ABI marginal costs by \$0.29 and Modelo marginal costs by \$1.80.

merger equilibrium satisfies the following first order conditions:

$$
\begin{aligned}
g_k(m^1, m^2; \eta) \;&\equiv\; \sum_{r=1}^{R} \left(\frac{1}{1-\eta} \pi_{kr}^{PLE}(m_r^1, m_r^2) - \pi_{kr}^{D,2}(m_r^1, m_r^2) - \frac{\eta}{1-\eta} \pi_{kr}^{B} \right) = 0 \\
h_r^1(m_r^1, m_R^1, m_r^2, m_R^2) \;&\equiv\; \frac{\frac{\partial \pi_{1r}^{PL}(m_r^1, m_r^2)}{\partial m_r^1}}{\frac{\partial \pi_{kr}^{D,k}(m_r^1, m_r^2)}{\partial m_r^1} - \frac{\partial \pi_{kr}^{PL}(m_r^1, m_r^2)}{\partial m_r^1}} - \frac{\frac{\partial \pi_{1R}^{PL}(m_R^1, m_R^2)}{\partial m_R^1}}{\frac{\partial \pi_{kR}^{D,k}(m_R^1, m_R^2)}{\partial m_{Rt}^1} - \frac{\partial \pi_{kR}^{PL}(m_R^1, m_R^2)}{\partial m_{Rt}^1}} = 0 \\
h_r^2(m_r^1, m_R^1, m_r^2, m_R^2) \;&\equiv\; \frac{\frac{\partial \pi_{1r}^{PL}(m_r^1, m_r^2)}{\partial m_r^2}}{\frac{\partial \pi_{kr}^{D,k}(m_r^1, m_r^2)}{\partial m_r^2} - \frac{\partial \pi_{kr}^{PL}(m_r^1, m_r^2)}{\partial m_r^2}} - \frac{\frac{\partial \pi_{1R}^{PL}(m_R^1, m_R^2)}{\partial m_R^1}}{\frac{\partial \pi_{kR}^{D,k}(m_R^1, m_R^2)}{\partial m_{Rt}^1} - \frac{\partial \pi_{kR}^{PL}(m_R^1, m_R^2)}{\partial m_{Rt}^1}} = 0
\end{aligned}
$$

An important detail is that the second ratios in the expressions for h^1 and h^2 are identical, and involve differentiating with respect to m_R^1. There are $R-1$ identifying equations for h^1 as, by inspection, the equation is always satisfied for $r = R$. However, there are R equations for h^2, so combining with the binding slack function (i.e., $g_k(m^1, m^2; \eta) = 0$) there are $2R$ equations in total, and the post-merger supermarkups are exactly identified.

6.3 Simulation Results

Table 6 summarizes the effects of the mergers on prices. The first two columns consider Miller/Coors with and without efficiencies. The final three columns consider ABI/Modelo under the three different efficiency scenarios. We report the average supermarkup change, the average Bertrand price change, and the total price change.[46]

Starting with Miller/Coors, we find that the merger increases the supermarkup by 0.51 in fiscal year 2010 (column (i)). Our imputation results indicate that the average supermarkup increases from 1.20 in fiscal year 2007 to 1.80 in fiscal year 2010 (Table 3). Thus, the Miller/Coors merger appears to account for 85% (0.51/0.60 = 0.85) of the change. We attribute the remainder to changing cost and demand conditions. Due to the merger efficiencies, Bertrand prices are largely unaffected, so the changes in total price largely reflect coordinated effects. Analyzing the merger without efficiencies, we find that the increase in the supermarkup decreases slightly, but the Bertrand price increases are higher, especially for Miller and Coors brands (column (ii)). Total prices are higher on average.

Turning to the ABI/Modelo merger, column (iii) of Table 6 shows that Bertrand prices

[46]We calculate the average supermarkups as means across the 37 regions, for comparability with Table 3. However, the average price changes are means across the product-region-quarter observations, so the average total price change is not always equal to the sum of the average supermarkup and Bertrand price changes.

Table 6: Merger Price Effects

Merger:	Miller/Coors		ABI/Modelo		
Efficiencies:	Observed	Removed	None	Minor	Major
	(i)	(ii)	(iii)	(iv)	(v)
Δ Supermarkup					
Domestic Brewers	0.51	0.39	0.40	0.37	0.32
Modelo	-	-	1.46	1.45	1.44
Δ Bertrand Price					
ABI	-0.01	0.16	0.20	0.21	0.00
MillerCoors	-	-	0.08	0.07	0.00
Miller	0.16	0.50	-	-	-
Coors	0.01	0.89	-	-	-
Modelo	-0.02	-0.04	1.84	1.31	0.00
Δ Total Price					
ABI	0.50	0.55	0.59	0.58	0.32
MillerCoors	-	-	0.48	0.44	0.32
Miller	0.67	0.90	-	-	-
Coors	0.52	1.28	-	-	-
Modelo	-0.03	-0.04	3.39	2.85	1.53

Notes: The table summarizes the price effects of mergers. Column (i) is a comparison of the observed equilibrium to a counterfactual without the Miller/Coors merger. Column (ii) is a comparison of a counterfactual in which the Miller/Coors merger occurs without efficiencies to a counterfactual in which the Miller/Coors merger does not occur. Columns (iii)-(v) are comparisons of counterfactuals in which an ABI/Modelo merger occurs, with varying levels of efficiencies, to the observed equilibrium. The Miller/Coors merger is evaluated in fiscal year 2010, whereas the ABI/Modelo merger is evaluated in fiscal year 2011. Statistics for the supermarkup change are averages across the 37 regions. Statistics for the changes in Bertrand prices, total prices, and market shares are averages over observations at the product-region-quarter level.

of ABI and Modelo products increase by \$0.20 and \$1.84, respectively, with the magnitude of the latter reflecting an incentive to steer customers toward higher-markup ABI brands. Prices also increase due to a higher supermarkup, which rises by \$0.40 for ABI and MillerCoors, and \$1.46 for Modelo. Efficiencies offset the unilateral incentive to raise prices, as in standard merger analysis, but in this case do little to reduce the supermarkup. This occurs because the impact of the marginal costs of ABI and Modelo on the supermarkup is indirect, coming through the (binding) MillerCoors slack function. The changes in total price reflect both the change in Bertrand price and the change in the supermarkup.

Table 7 decomposes the binding slack function across various scenarios to explore in greater detail how the mergers impact coordination incentives. Columns (i) and (iv) contain pre-merger values for the Miller/Coors and ABI/Modelo mergers, respectively. Coors is the binding firm prior to the Miller/Coors merger, while MillerCoors is the binding firm prior to the ABI/Modelo merger.

Columns (ii) and (v) present the slack function and its components after each merger but at pre-merger supermarkups and costs. With Miller/Coors, the slack function changes due to the unilateral effect on Bertrand prices and because the binding post-merger slack function incorporates Miller products. With ABI/Modelo, the slack function changes only due to the effect on Bertrand prices. In both mergers, row (e) shows that the present value of deviating increases as both punishment phase competition is softened (row (c)) and profits from deviating increase (row (b)). This mechanism has been highlighted in several empirical and theoretical papers that show mergers can make tacit collusion more difficult when collusion is at joint monopoly prices (e.g., Davidson and Deneckere (1984); Werden and Baumann (1986); Davis and Huse (2010)). However, in our model collusion is not at the joint monopoly level, so even a merger between tacitly colluding firms can make collusion easier. Indeed, each merger increases price leadership profits by more than deviation profits, allowing for higher supermarkups.

Columns (iii) and (vii) illustrate how efficiencies can change incentives to collude. In the case of Miller/Coors, efficiencies increase supermarkups. Comparing column (iii) to column (ii) shows that while efficiencies increase each component of the slack function, the value of price leadership increases by more than the total deviation value. This allows for greater supermarkups. In contrast, comparing columns (vi) and (vii) demonstrates that efficiencies very slightly reduce the slack function and as a result supermarkups fall in the case of ABI/Modelo. (Here, we focus on the minor efficiencies scenario.) Together, these simulation results indicate that efficiencies have ambiguous implications for the magnitude of coordinated effects.

Table 8 provides welfare effects for the two mergers. As a result of Miller/Coors, industry profits increase by 11.4% with the merging firms experiencing the largest increase because of efficiencies. Consumer surplus falls by \$1.06 for every dollar gained in industry profits. Turning to ABI/Modelo, industry profits increase by between 9.95% and 10% depending on whether major efficiencies are incorporated. Even with major efficiencies, consumers lose \$1.59 in surplus for every dollar gained by firms.

Table 7: Decomposition of Binding Slack Functions

		Miller/Coors Merger				ABI/Modelo Merger			
Ownership:		Pre	Post	Post	Post	Pre	Post	Post	Post
Efficiencies:		No	No	Yes	Yes	No	No	Yes	Yes
Supermarkup:		Pre	Pre	Pre	Post	Pre	Pre	Pre	Post
		(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)
(a)	π^{PL}	11.85	35.08	41.12	42.78	39.70	42.85	42.67	43.73
(b)	π^{D}	12.77	36.45	42.53	45.44	42.15	45.06	44.88	47.06
(c)	π^{B}	9.22	29.84	35.22	35.22	32.72	34.41	34.24	34.24
(d)	$\frac{1}{1-\eta}\pi^{PL}$	16.01	47.41	55.57	57.82	53.65	57.90	57.66	59.09
(e)	$\pi^{D}+\frac{\eta}{1-\eta}\pi^{B}$	16.01	46.93	54.90	57.82	53.65	57.15	56.91	59.09
(f)	$g(m)$	0	0.474	0.668	0	0	0.749	0.747	0

Notes: The table shows the components of the binding slack function for Miller/Coors and ABI/Modelo. Units are millions of dollars. Slack functions are calculated using the timing parameter $\eta = 0.26$. The Miller/Coors merger is evaluated in fiscal year 2010, and the ABI/Modelo merger is evaluated in fiscal year 2011. The profit statistics are at the firm- or industry-level. Column (iii) adds estimated efficiencies resulting from Miller/Coors while column (vii) reduces Modelo costs by \$0.50.

7 Conclusion

This study represents an attempt to apply methodologies that have become standard in the industrial organization literature over the previous two decades to a repeated pricing game of perfect information. The particular setting—price leadership in the beer industry—is advantageous in part because documentary evidence in the public record informs the timing of actions and the strategies that firms play along the equilibrium path. However, the methods we use could be adapted to other repeated pricing games of perfect information, including but not limited to other markets which exhibit price leadership. The general insight is that the empirical modeling of collusion and supracompetitive pricing is feasible if one has sufficient knowledge of the underlying game.

One practical benefit of our modeling approach is that it allows for the prospective evaluation of mergers when coordinated effects are a concern. In our context, we use counterfactual simulations to demonstrate that pricing can become less competitive due to (i) a merger that absorbs the firm with a binding IC constraint into a larger entity, and (ii) a merger between a coalition member and a competitive fringe firm. In both of these cases, even substantial merger efficiencies are insufficient to mitigate price effects. One limitation of our approach is that it requires coordination to be present in the pre-merger data because this allows for IC constraints to be evaluated, and then adjusted for post-merger

Table 8: Mergers Effects on Profit and Consumer Surplus

Merger:	Miller/Coors		ABI/Modelo		
Efficiencies:	Observed	Removed	None	Minor	Major
	(i)	(ii)	(iii)	(iv)	(v)
% Δ Profit					
All Firms	11.39	8.38	9.95	9.95	11.72
ABI	3.24	9.72	-	-	-
ABI+Modelo	-	-	7.25	7.99	15.54
Miller+Coors	22.58	4.48	11.55	10.81	6.36
Modelo	8.86	16.71	-	-	-
Δ CS / Δ Profit	-1.06	-1.95	-2.31	-2.24	-1.59

Notes: The table summarizes the effects of mergers on profit and consumer surplus. Column (i) is a comparison of the observed equilibrium to a counterfactual in which the Miller/Coors merger does not occur. Column (ii) is a comparison of a counterfactual in which the Miller/Coors merger occurs without efficiencies to a counterfactual in which the Miller/Coors merger does not occur. Columns (iii)-(v) are comparisons of counterfactuals in which an ABI/Modelo merger occurs, at varying levels of efficiencies, to the observed equilibrium. The Miller/Coors merger is evaluated in fiscal year 2010, and the ABI/Modelo merger is evaluated in fiscal year 2011. The profit statistics are at the firm- or industry-level. The consumer surplus statistics are calculated by aggregating over regions and quarters, and then normalizing by the change in total industry profit.

conditions. The model does not predict when a coordinated equilibrium would emerge due to a merger. It is also worth noting that our approach to merger simulation holds the timing parameter—and thus the duration of deviation and punishment—fixed. To the extent timing considerations are endogenously determined by market structure, a version of the Lucas (1976) critique applies. Additionally, our counterfactuals hold fixed the identities of the coalition members. A way to relax this assumption might be to simulate counterfactual profits for each potential coalition. Then a rule for cartel formation, for example that in Bos and Harrington (2010), could be used to determine which are stable.

We view our model as an empirical version of two canonical models of collusion with perfect information (Rotemberg and Saloner (1986); Bernheim and Whinston (1990)). Some phenomena, such as price wars or the signaling of private information, do not occur. Thus, we believe that future research extending our methods to empirical settings of imperfect information, such as those examined theoretically in Green and Porter (1984) and Rotemberg and Saloner (1990), would be a useful complement to our work.

References

Abreu, D. (1986). Optimal cartel equilibrium with imperfect monitoring. *Journal of Economic Theory*.

Ascher, B. (2012). Global beer: The road to monopoly. American Antitrust Institute monograph.

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

Asker, J. (2010). A study of the internal organization of a bidding cartel. *The American Economic Review 100*(3), 724–762.

Asker, J. (2016). Diagnosing Foreclosure due to Exclusive Dealing. *Journal of Industrial Economics 64*(3), 375–410.

Asker, J., A. Collard-Wexler, and J. D. Loecker (2019). (Mis)Allocation, market power, and global oil extraction. *American Economic Review 109*(4), 1568–1615.

Aumann, R. (1990). Nash equilibrium are not self-enforcing. In Gabszewicz, J. J., J.-F. Richard, and L. A. Wolsey, eds., *Economic Decision-Making: Games, Econometrics and Optimisation*. Amsterdam: Elsevier, pp. 201-206.

Bain, J. S. (1960). Price leaders, barometers, and kinks. *Journal of Business 33*(3), 193–203.

Baker, J. B. (2001). Mavericks, mergers, and exclusion: Proving coordinated competitive effects under the antitrust laws. *New York University Law Review 77*, 135–203.

Baker, J. B. (2010). Market concentration in the antitrust analysis of horizontal mergers.

Bernheim, B. D. and M. D. Whinston (1990). Multimarket contact and collusive behavior. *RAND Journal of Economics 21*(1), 1–26.

Berry, S., J. Levinsohn, and A. Pakes (1995, July). Automobile prices in market equilibrium. *Econometrica 63*(4), 847–890.

Bos, I. and J. E. Harrington (2010). Endogenous Cartel Formation with Heterogeneous Firms. *RAND Journal of Economics 41*(1), 92–117.

Bresnahan, T. (1987). Competition and collusion in the american automobile oligopoly: The 1955 price war. *Journal of Industrial Economics 35*(4), 457–482.

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

Butters, R. A., D. W. Sacks, and B. Seo (2020). How do national firms respond to local shocks? evidence from excise taxes.

Byrne, D. P. and N. de Roos (2019). Learning to Coordinate: A Study in Retail Gasoline. *American Economic Review 109*(2), 591–619.

Chilet, J. A. (2018). Gradually rebuilding a relationship: The emergence of collusion in retail pharmacies in Chile.

Ciliberto, F. and J. W. Williams (2014). Does Multimarket Contact Facilitate Tacit Collusion? Inference on Conduct Parameters in the Airline Industry. *RAND Journal of Economics 45*(4), 764–791.

Clark, R. and J.-F. Houde (2013, August). Collusion with asymmetric retailers: Evidence from a gasoline price-fixing case. *American Economic Journal: Microeconomics 5*(3), 97–123.

Compte, O., F. Jenny, and P. Rey (2002). Capacity constraints, mergers and collusion. *European Economic Review 46*(1), 1–29.

d'Aspremont, C., A. Jacquemin, J. J. Babszewicz, and J. A. Weymark (1983). On the Stability of Collusive Price Leadership. *Canadian Journal of Economics 26*, 17–25.

Davidson, C. and R. Deneckere (1984). Horizontal mergers and collusive behavior. *International Journal of Industrial Organization 2*(2), 117–132.

Davis, P. J. and C. Huse (2010). Estimating the "coordinated effects" of mergers. *Working Paper*.

DellaVigna, S. and M. Gentzkow (2019). Uniform pricing in U.S. retail chains? *Quarterly Journal of Economics 134*(4), 2011–2084.

Donsimoni, M., N. Economides, and H. Polemarchakis (1986). Stable Cartels. *International Economic Review 27*(2), pp. 317–327.

Drennan, R., M. Magura, and A. Nevo (2013). The year in review: Economics at the Antitrust Division 2012-2013. *Review of Industrial Organization 43*(4), 291–302.

Eizenberg, A. and D. Shilian (2019). Structure, conduct, and contact: Competition in closely-related markets. *mimeo*.

Eizenberg, A., D. Shilian, and D. D. Blanga (2020). Estimating the potential impact of mulitmarket contact on the instensity of competition. *mimeo*.

Evans and Kessides (1994). Living by the golden rule: Multimarket contact in the US airline industry. *Journal of Quarterly Economics 109*(2), 341–366.

Farrell, J. and M. Rabin (1996). Cheap talk. *Journal of Economic Perspectives 10*(3), 103–118.

Gasmi, F., J.-J. Laffont, and Q. Vuong (1992). Econometric analysis of collusive behavior in a soft-drink market. *Journal of Economics & Management Strategy 1*(2), 277–311.

Green, E. and R. Porter (1984). Noncooperative collusion under imperfect price information. *Econometrica 52*(1), 87–100.

Greer, D. F. (1998). Beer: Causes of structural change. In *Industry Studies*, 2nd Edition, Larry L. Duetsch (ed), M.E. Sharpe, New York.

Grigolon, L. and F. Verboven (2014). Nested logit or random coefficient logit? A comparison of alternative discrete choice models of product differentiation. *Review of Economics and Statistics 96*(5), 916–935.

Harrington, J. E. (2013). Evaluating mergers for coordinated effects and the role of "parallel accommodating conduct". *Antitrust Law Journal* (78), 651–668.

Harrington, J. E., R. H. Gonzalez, and P. Kujal (2016). The relative efficacy of price announcements and express communication for collusion: Experimental findings. *Journal of Economic Behavior & Organization 128*, 251–264.

Harrington, J. E. and P. T. Harker (2018). Developing competition law for collusion by autonomous price-setting agents. *Journal of Competition Law and Economics 14*, 331–363.

Harrington, J. E. and A. Skrzypacz (2011). Private monitoring and communication in cartels: Explaining recent collusive practices. *American Economic Review 101*(6), 2425–2449.

Igami, M. (2015). Market Power in International Commodity Trade: The case of coffee. *Journal of Industrial Economics 63*(2), 225–248.

Igami, M. and T. Sugaya (2020). Measuring the incentive to collude: The vitamin cartels, 1990-1999. Available at SSRN: `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2889837`.

Ishibashi, I. (2008). Collusive price leadership with capacity constraints. *International Journal of Industrial Organization 26*(3), 704–715.

Ivaldi, M., B. Jullien, P. Rey, P. Seabright, and J. Tirole (2007). The economics of tacit collusion: Implications for merger control. In *The Political Economy of Antitrust,* Vivek Ghosal, and Johan Stennek (eds.), Elsevier Science, 217-240.

Lanzillotti, R. F. (2017). Collusion/Competition: A New Learning? *Antitrust Bulletin 62*(3), 591–602.

Lemus, J. and F. Luco (2018). Pricing dynamics and leadership: Evidence from the retail gasoline industry. Available at SSRN: `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3186144`.

Loertscher, S. and L. M. Marx (2019). Merger review for markets with buyer power. *Journal of Political Economy 127*(6), 2967–3017.

Loertscher, S. and L. M. Marx (2020). Coordinated effects. *mimeo*.

Markham, J. W. (1951). The nature and significance of price leadership. *Americal Economic Review 41*(5), 891–905.

McClain, J. (2012). 2011-2012 annual report. Technical report, Beer Institute, Chicago Illinois.

McFadden, D. (1978). Modelling of choice of residential location. In *Spatial Interaction Theory and Residential Location*, F. Snickers, A. Karlquist, L. Lundquist, J. Weibull (eds), Amsterdam: North Holland.

Michel, C. and S. Weiergraeber (2018). Estimating industry conduct in differentiated products markets: The evolution of pricing behavior in the RTE cereal industry. *mimeo*.

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

Mouraviev, I. and P. Rey (2011). Collusion and leadership. *International Journal of Industrial Organization 29*, 705–717.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica 69*(2), pp. 307–342.

Nocke, V. and N. Schutz (2018). Multiproduct-firm oligopoly: An aggregative games approach. *Econometrica 86*(2), 523–557.

Oxenfeldt, A. R. (1952). Professor Markham on price leaderhip: Some unanswered questions. *American Economic Review 42*(3), 380–384.

Pastine, I. and T. Pastine (2004). Cost of delay and endogenous price leadership. *International Journal of Industrial Organization 22*(1), 135–145.

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

Rojas, C. (2008). Price competition in U.S. brewing. *Journal of Industrial Economics 66*(1), 1–31.

Rosse, J. N. (1970). Estimating cost function parameters without using cost data: Illustrated methodology. *Econometrica 38*(2), 256–275.

Rotemberg, J. J. and G. Saloner (1986). A supergame-theoretical model of price wars during booms. *American Economic Review 76*(3), 390–407.

Rotemberg, J. J. and G. Saloner (1990). Collusive price leadership. *Journal of Industrial Economics 39*(1), 93–111.

Scherer, F. (1980). *Industrial Market Structure and Economic Performance*. Houghton Mifflin.

Seaton, J. S. and M. Waterson (2013). Identifying and characterising price leadership in British supermarkets. *International Journal of Industrial Organization 31*, 392–403.

Shim, B. and A. Khwaja (2017). The collusive effect of multimarket contact on prices: Evidence from retail lumber markets. 2017 Meeting Papers 593, Society for Economic Dynamics.

Slade, M. E. (2004). Market power and joint dominance in U.K. brewing. *Journal of Industrial Economics 52*(1), 133–163.

Stigler, G. J. (1947). The kinky demand curve and rigid prices. *Journal of Political Economy 55*(5).

Tremblay, V. J. and C. H. Tremblay (2005). *The U.S. Brewing Industry.* The MIT Press.

Vasconcelso, H. (2005). Tacit collusion, cost asymmetries, and mergers. *RAND Journal of Economics 36*(1), 39–62.

Werden, G. J. (1996). A robust test for consumer welfare enhancing mergers among sellers of differentiated products. *The Journal of Industrial Economics 44*(4), pp. 409–413.

Werden, G. J. and M. G. Baumann (1986). A simple model of imperfect competition in which four are few but three are not. *Journal of Industrial Economics 34*(3), 331–335.

Appendix

A Data

The brands in our sample include Bud Light, Budweiser, Michelob, Michelob Light, Miller Lite, Miller Genuine Draft, Miller High Life, Coors Light, Coors, Corona Extra, Corona Extra Light, Heineken, and Heineken Light. Miller and Weinberg (2017) report that these brands account for 68% of all unit sales of SAB Miller, Molson Coors, ABI, Modelo, and Heineken. The most popular brands that are excluded are regional brands (e.g., Yuengling Lager, Labatt Blue) or subpremium brands that sell at lower price points (e.g., Busch Light, Natural Light, Busch, Keystone, Natural Ice). Many of the subpremium brands are owned by ABI. Also excluded are some brands that enter or exit during the sample period (e.g., Budweiser Select, Bud Light Lime). Adding brands to the model would require re-estimation of the RCNL demand model.

We restrict our attention to 6 packs, 12 packs, 24 packs, and 30 packs, the latter two of which we combine in the construction of our products. Miller and Weinberg (2017) report that these sizes account for 75% of all unit sales among the brands that we consider.

We use the demographic draws of Miller and Weinberg (2017), which are from the Public Use Microdata Sample (PUMS) of the American Community Survey, in the demand model. The PUMS data are available annually over 2005–2011. Households are identified as residing within specified geographic areas, each of which has at least 100,000 residents based on the 2000 U.S. Census. The PUMS data are merged with the IRI scanner data by matching on the counties that compose the IRI regions and the PUMS areas. There are 500 draws on households per region–year. Income is measured as total household income divided by the number of household members.

Miller and Weinberg (2017) restrict attention to 39 of the 47 geographic regions in the IRI academic database, dropping a handful of regions in which either few supermarkets are licensed to sell beer or supermarkets are restricted to selling low-alcohol beer.[47] The IRI data are not intended to be fully representative of regions in the United States. Bronnenberg et al. (2008, p. 746) states that

[47] The regions included in the Miller and Weinberg (2017) sample are Atlanta, Birmingham/Montgomery, Boston, Buffalo/Rochester, Charlotte, Chicago, Cleveland, Dallas, Des Moines, Detroit, Grand Rapids, Green Bay, Hartford, Houston, Indianapolis, Knoxville, Los Angeles, Milwaukee, Mississippi, New Orleans, New York, Omaha, Peoria/Springfield, Phoenix, Portland in Oregon, Raleigh/Durham, Richmond/Norfolk, Roanoke, Sacramento, San Diego, San Francisco, Seattle/Tacoma, South Carolina, Spokane, St. Louis, Syracuse, Toledo, Washington D.C., and West Texas/New Mexico.

> In practice, to protect the confidentiality of these chains, markets in which the top grocery chain has more than 50% of the grocery market are omitted. This reduces the coverage to 47 markets.

If the excluded regions feature more buyer power than those included in the sample, which seems plausible given this inclusion rule, then the supermarkups we find might exceed the national average. See the discussion in Section 3.6 for more on buyer power.

In our sample, we also exclude New Orleans and San Diego, leaving a total of 37 geographic regions in the data. The exclusion is due to a computational issue related to the demographic draws. When we compute the fringe's best responses to even very small supermarkups, such as $m = 0.01$, we find that Modelo and Heineken respond by *lowering* prices in New Orleans and San Diego, often by an order of magnitude more than the supermarkup. The magnitude and direction of this response—and note that prices are acting as strategic substitutes here—makes any positive supermarkup unprofitable. We attribute this outsize strategic substitutes response to poorly-performing demographic draws. In numerical checks based on logit demand, which omits the impact of the demographic draws, we always compute PLE that feature positive supermarkups (Appendix C.3). Consistent with this, the income draws of New Orleans and San Diego are much larger on average than nearly all other regions, and also exhibit greater variance.[48] Thus, we believe the computational issue is due to the demand-side of the model and is not indicative of a broader supply-side performance issue.

B Theoretical Matters

B.1 Slack Functions and the Timing Parameter

Here we provide a proof to Proposition 2. We show that a slack function with (1) deviation profits being earned for one or more periods, (2) punishment occurring for a finite period of time, and (3) some probability that the market ceases to exist can be restated in an equivalent form with (i) deviation profits being earned for one period, (ii) punishment occurring by grim trigger, and (iii) the market continuing to exist each period with probability one. Formally,

[48] Variance is relevant because prices tend to be strategic substitutes only with enough consumer heterogeneity (or "demand curvature"). Miller and Weinberg (2017, Table A.2 in the Online Appendix) shows that New Orleans and San Diego each have HHIs that are comparable to other regions in the data.

assumptions (1) - (3) give a region-specific contribution to the slack function of

$$\begin{aligned}\tilde{g}_{irt}(m_{rt};\cdot) \;&\equiv\; \left(\pi^{PL}_{irt}(m_{rt};\Psi_t)+\frac{\phi\delta}{1-\phi\delta}\pi^{PL}_{irt}(\tilde{m}_{rt};\Psi_t)\right)\\ &\quad -\left(\sum_{s=0}^{\tau_1-1}(\phi\delta)^s\pi^{D,i}_{irt}(m_{rt};\Psi_t)+\sum_{s=\tau_1}^{\tau_1+\tau_2-1}(\phi\delta)^s\pi^{B}_{irt}(;\Psi_t)+\sum_{s=\tau_1+\tau_2}^{\infty}(\phi\delta)^s\pi^{PL}_{irt}(\tilde{m}_{rt};\Psi_t)\right),\end{aligned}$$

which is equation (3). Because $\phi \le 1$ and $\delta < 1$, we have $|\phi\delta| < 1$. Thus, the geometric series can be simplified, giving

$$\begin{aligned}\tilde{g}_{irt}(m_{rt};\cdot) \;&\equiv\; \frac{1-(\phi\delta)^{\tau_1}}{1-\phi\delta}\Bigg(\left(\pi^{PL}_{irt}(m_{rt};\Psi_t)+\frac{(\phi\delta)^{\tau_1}-(\phi\delta)^{\tau_1+\tau_2}}{1-(\phi\delta)^{\tau_1}}\pi^{PL}_{irt}(\tilde{m}_{rt};\Psi_t)\right)\\ &\quad -\left(\pi^{D,i}_{irt}(m_{rt};\Psi_t)+\frac{(\phi\delta)^{\tau_1}-(\phi\delta)^{\tau_1+\tau_2}}{1-(\phi\delta)^{\tau_1}}\pi^{B}_{irt}(;\Psi_t)\right)\Bigg).\end{aligned}$$

If we set η such that

$$\eta=\frac{(\phi\delta)^{\tau_1}-(\phi\delta)^{\tau_1+\tau_2}}{1-(\phi\delta)^{\tau_1+\tau_2}},$$

then we see that $\tilde{g}_{irt} = ((1-(\phi\delta)^{\tau_1})/(1-\phi\delta))g_{irt}$. Let $\psi = (1-(\phi\delta)^{\tau_1})/(1-\phi\delta)$. Summing across regions, this also implies that $\tilde{g}_{it} = \psi g_{it}$. Because $|\phi\delta| < 1$, we have $\psi > 0$. Furthermore, if $\tau_1 = 1$, then $\psi = 1$.

QED

B.2 The Leader's Constrained Maximization Problem

In this appendix, we derive the first order conditions that characterize the solution to the leader's constrained maximization problem. The Lagrangian is given by

$$L=\sum_{r\in\mathbb{R}}\pi^{PL}_{1rt}(m_{rt};\Psi_t)+\sum_{j\in\mathbb{C}}\lambda_j g_{jt}(m_t;\eta,\tilde{m}_t,\Psi_t),$$

where λ_j is the Lagrange multiplier for the IC constraint of firm j. This gives a series of first order conditions of the form

$$\frac{\partial L}{\partial m_{rt}}=\frac{\partial \pi^{PL}_{1rt}(m_{rt};\Psi_t)}{\partial m_{rt}}+\lambda_k\left(\frac{\partial g_{kt}(m_t;\eta,\tilde{m}_t,\Psi_t)}{\partial m_{rt}}\right)=0\quad \forall r\in\mathbb{R},$$

where firm k has the IC constraint that binds. The derivative of the slack function for coalition member k with respect to the supermarkup for region r is

$$\frac{\partial g_{kt}(m_t;\eta,\tilde{m}_t,\Psi_t)}{\partial m_{rt}} = -\frac{\partial \pi^{D,k}_{krt}(m_{rt};\Psi_t)}{\partial m_{rt}} + \frac{\partial \pi^{PL}_{krt}(m_{rt};\Psi_t)}{\partial m_{rt}}.$$

Plugging in for this derivative and then rearranging the first order condition for region R gives

$$\lambda_k = \frac{\frac{\partial \pi^{PL}_{1Rt}(m_{Rt};\Psi_t)}{\partial m_{Rt}}}{\frac{\partial \pi^{D,k}_{kRt}(m_{Rt};\Psi_t)}{\partial m_{Rt}} - \frac{\partial \pi^{PL}_{kRt}(m_{Rt};\Psi_t)}{\partial m_{Rt}}}.$$

Substituting into the other first order conditions gives

$$\frac{\frac{\partial \pi_{1rt}(m_{rt};\Psi_t)}{\partial m_{rt}}}{\frac{\partial \pi^{D,k}_{krt}(m_{rt};\Psi_t)}{\partial m_{rt}} - \frac{\partial \pi^{PL}_{krt}(m_{rt};\Psi_t)}{\partial m_{rt}}} = \frac{\frac{\partial \pi^{PL}_{1Rt}(m_{Rt};\Psi_t)}{\partial m_{Rt}}}{\frac{\partial \pi^{D,k}_{kRt}(m_{Rt};\Psi_t)}{\partial m_{Rt}} - \frac{\partial \pi^{PL}_{kRt}(m_{Rt};\Psi_t)}{\partial m_{Rt}}}$$

for all other regions $r \in \mathbb{R} \setminus \{1\}$.

C Numerical Matters

In this appendix we describe the methods we use for imputation and simulation, under the assumption that prices are generated by a constrained PLE. We assume knowledge of the timing parameter, η. We focus mainly on the baseline model, which features a single supermarkup for each region and IC constraints that pool across regions. We also discuss alternative models with (i) separate supermarkups for small and large product sizes, and (ii) IC constraints that do not pool across regions. For the purposes of this appendix section, assume that firm 1 is the leader and that firm k has the binding constraint. This is without loss of generality—the leader and the binding firm could be any of the coalition firms, and could even be the same firm. We also drop time subscripts for notational brevity, as neither imputation nor simulation require multiple periods.

C.1 Imputation with the Baseline Model

We start with the case in which the leader selects a single supermarkup for each region. The solution to the leader's constrained maximization problem requires that the vector of supermarkups, $m = (m_1, m_2, \ldots, m_R)$, satisfies the following equations:

$$g_k(m;\eta) \equiv \sum_{r=1}^{R} \left(\frac{1}{1-\eta} \pi_{kr}^{PL}(m_r) - \pi_{kr}^{D,k}(m_r) - \frac{\eta}{1-\eta} \pi_{kr}^{B} \right) = 0 \tag{C.1}$$

$$h_r(m_r, m_R) \equiv \frac{\frac{\partial \pi_{1r}^{PL}(m_r)}{\partial m_r}}{\frac{\partial \pi_{kr}^{D,k}(m_r)}{\partial m_r} - \frac{\partial \pi_{kr}^{PL}(m_r)}{\partial m_r}} - \frac{\frac{\partial \pi_{1R}^{PL}(m_R)}{\partial m_R}}{\frac{\partial \pi_{kR}^{D,k}(m_{Rt})}{\partial m_{Rt}} - \frac{\partial \pi_{kR}^{PL}(m_{Rt})}{\partial m_{Rt}}} = 0 \tag{C.2}$$

where (C.2) applies to regions $r = 1, \ldots, R-1$. Equation (C.1) can be obtained from equations (5) and (6) after applying the rational expectations assumption. This is a system with R nonlinear equations and R unknowns.

In principle, a solution could be obtained using off-the-shelf algorithms, such as an equation solver (e.g., fsolve in Matlab) or a minimizer (e.g., fminsearch in Matlab). However, that approach appears to be computationally prohibitive in our application, for two reasons. The first is that solving a nonlinear system of equations is inherently slow if there are many parameters. We have 37 regions, and thus 37 parameters, which is enough to make the search difficult even with fast function evaluations. The second reason is that function evaluation is quite slow. For any candidate $\tilde{m}$, Proposition 1 must be applied to recover marginal costs and Bertrand prices. Then, with marginal costs in hand, deviation prices must be computed. Finally, the derivatives in (C.2) must be obtained numerically, so many of the steps must be repeated. Even with streamlined code and using multiple processors, we have been unable to make substantial progress with this approach.

Instead, we develop an approach to imputation that better exploits the economics of the model to obtain solutions. Notice that any candidate parameter vector can be expressed $\widetilde{m} = (\widetilde{m}_R, \widetilde{\Delta m})$, where $\widetilde{\Delta m} = (\widetilde{\Delta m}_1, \ldots, \widetilde{\Delta m}_{R-1})$ and $\widetilde{\Delta m}_r \equiv \widetilde{m}_r - \widetilde{m}_R$. The vector $\widetilde{m}$ thus contains the supermarkup that applies to region R and the supermarkup differences between region R and the other regions. Our strategy involves solving for $\widetilde{m}_R$ holding fixed $\widetilde{\Delta m}$, then solving for $\widetilde{\Delta m}$ holding fixed $\widetilde{m}_R$. Repeating this multiple times, we are able to reliably find a solution that satisfies all the first order conditions to any arbitrary degree of accuracy. Specifically, the steps are:

1. Set $\widetilde{\Delta m}^{(0)} = \overrightarrow{0}$. Search for the $\widetilde{m}_R$ that solves (C.1). We use fminsearch in Matlab with a tolerance of 1e-4.[49] Each evaluation of the loss function requires a number of numerical steps, so the search is nontrivial. However, it is feasible nonetheless because

[49]The tolerance is not unit-free, but given the slack function values we obtain away from the optimum (e.g., Figure 4), the tolerance of 1e-4 is quite tight. Using an even tighter tolerance does not affect results.

there is only one unknown. At the solution, the IC constraint binds, but the dispersion of supermarkups across regions is sub-optimal. Denote the solution $\widetilde{m}_R^{(0)}\left(\widetilde{\Delta m}^{(0)}\right)$.

2. Find the $\widetilde{\Delta m}$ that solves (C.2), holding fixed $\widetilde{m}_R^{(0)}\left(\widetilde{\Delta m}^{(0)}\right)$. At the solution, the dispersion of supermarkups across regions is optimal but an IC may no longer bind. Denote the solution $\widetilde{\Delta m}^{(1)}\left(\widetilde{m}_R^{(0)}\right)$. The direct approach of solving for the $R-1$ supermarkup differences based on the $R-1$ nonlinear equations (e.g., with Matlab's fminsearch) is computationally prohibitive. However, consider the RHS of (C.2):

$$\frac{\frac{\partial \pi_{1r}^{PL}(m_r)}{\partial m_r}}{\frac{\partial \pi_{kr}^{D,k}(m_r)}{\partial m_r}-\frac{\partial \pi_{kr}^{PL}(m_r)}{\partial m_r}}-\frac{\frac{\partial \pi_{1R}^{PL}(m_R)}{\partial m_R}}{\frac{\partial \pi_{kR}^{D,k}(m_{Rt})}{\partial m_{Rt}}-\frac{\partial \pi_{kR}^{PL}(m_{Rt})}{\partial m_{Rt}}}$$

The second ratio is invariant to $\widetilde{\Delta m}$, which does not affect profit in region R. The denominator of the first ratio increases in $m_r = m_R + \Delta m_r$ because increasing the supermarkup raises deviation profit more than price leadership profit. Similarly, the numerator of the first ratio decreases with m_r because of diminishing returns to increasing the supermarkup. Thus, the value of $h(m_r)$ can be adjusted up or down in predictable ways by changing $\widetilde{\Delta m}_r$. We apply the following sub-routine:

(a) Set $i=1$. Obtain h_r^i from (C.2), evaluating at $\widetilde{\Delta m}^i = \widetilde{\Delta m}^{(0)}$ and $\widetilde{m}_R^{(0)}\left(\widetilde{\Delta m}^{(0)}\right)$, and do so for each $r=1,\ldots,R-1$.

(b) Update $\widetilde{\Delta m}_r^{i+1} = \widetilde{\Delta m}_r^{i} + \phi \times \text{sign}\left(h^i\right)$ for some sufficiently small step ϕ.

(c) Obtain h_r^{i+1}, evaluating at $\widetilde{\Delta m}_r^{i+1}$ and $\widetilde{m}_R^{(0)}\left(\widetilde{\Delta m}^{(0)}\right)$, for each $r=1,\ldots,R-1$.

(d) Repeat (b) and (c) until a convergence criterion is satisfied. One option is to formulate the criterion as $\max_{r=1,\ldots,R-1}(|h_r|) < \epsilon$, for some small ϵ.

This reliably obtains the solution if the step size ϕ is not too large. The algorithm makes imputation feasible because a single function evaluation allows for all of the parameters to be adjusted simultaneously, and we find that computational time scales roughly linearly with the number of supermarkups. However, more work is needed because (C.1) may no longer hold.

3. Search for the $\widetilde{m}_R$ that solves (C.1), holding fixed $\widetilde{\Delta m}^{(j)}\left(\widetilde{m}_R^{(j-1)}\right)$, as in step 1. Denote the solution $\widetilde{m}_R^{(j)}\left(\widetilde{\Delta m}^{(j)}\right)$.

4. Find the $\widetilde{\Delta m}$ that solves (C.2), holding fixed $\widetilde{m}_R^{(j)}\left(\widetilde{\Delta m}^{(j-1)}\right)$, as in step 2. Denote the solution $\widetilde{\Delta m}^{(j+1)}\left(\widetilde{m}_R^{(j)}\right)$.

5. Iterate on steps 3 and 4, using progressively tighter tolerances in step 4. We iterate eight times. The final tolerance is such that no h_r is more than 0.0001 percent different than h_R.

6. Repeat step 3.

The final step prioritizes (C.1) over (C.2) in the imputation routine. However, the iterations in step 5 ensure that it does not substantially affect whether (C.2) is satisfied. In our baseline imputations, no h_r is more than 0.001 percent different than h_R. This imputation procedure can be parallelized at various point to take advantage of multiple processors. With the computer resources we have used, implementation has been feasible, but slow. Given that we use 500 consumer draws in the RCNL demand model, we expect that implementation with logit or nested logit demand models—which do not require numerically integrating over consumer draws—would be approximately 500 times faster.

C.2 Imputation with Alternative Assumptions

Size-Specific Supermarkups

We now consider the case in which the leader sets two supermarkups: one that applies to 6 and 12 packs, and another that applies to 24 packs. There are $2R$ supermarkups that must be recovered. Denote the supermarkups for the 6 and 12 packs as $m^1 = (m_1^1, m_2^1, \ldots, m_R^1)$ and the markups for the 24 packs as $m^2 = (m_1^2, m_2^2, \ldots, m_R^2)$. An evaluation of the leader's Lagrangian provides the following first order conditions:

$$\begin{aligned}
g_k(m^1, m^2; \eta) &\equiv \sum_{r=1}^{R}\left(\frac{1}{1-\eta}\pi_{kr}^{PL}(m_r^1, m_r^2) - \pi_{kr}^{D,k}(m_r^1, m_r^2) - \frac{\eta}{1-\eta}\pi_{kr}^{B}\right) = 0 \\
h_r^1(m_r^1, m_R^1, m_r^2, m_R^2) &\equiv \frac{\frac{\partial \pi_{1r}^{PL}(m_r^1, m_r^2)}{\partial m_r^1}}{\frac{\partial \pi_{kr}^{D,k}(m_r^1, m_r^2)}{\partial m_r^1} - \frac{\partial \pi_{kr}^{PL}(m_r^1, m_r^2)}{\partial m_r^1}} - \frac{\frac{\partial \pi_{1R}^{PL}(m_R^1, m_R^2)}{\partial m_R^1}}{\frac{\partial \pi_{kR}^{D,k}(m_R^1, m_R^2)}{\partial m_{Rt}^1} - \frac{\partial \pi_{kR}^{PL}(m_R^1, m_R^2)}{\partial m_{Rt}^1}} = 0 \\
h_r^2(m_r^1, m_R^1, m_r^2, m_R^2) &\equiv \frac{\frac{\partial \pi_{1r}^{PL}(m_r^1, m_r^2)}{\partial m_r^2}}{\frac{\partial \pi_{kr}^{D,k}(m_r^1, m_r^2)}{\partial m_r^2} - \frac{\partial \pi_{kr}^{PL}(m_r^1, m_r^2)}{\partial m_r^2}} - \frac{\frac{\partial \pi_{1R}^{PL}(m_R^1, m_R^2)}{\partial m_R^1}}{\frac{\partial \pi_{kR}^{D,k}(m_R^1, m_R^2)}{\partial m_{Rt}^1} - \frac{\partial \pi_{kR}^{PL}(m_R^1, m_R^2)}{\partial m_{Rt}^1}} = 0
\end{aligned}$$

An important detail is that the second ratios in the expressions for h^1 and h^2 are identical, and involve differentiating with respect to m_R^1. Thus, there are $R-1$ equations available for h^1 but R equations available for h^2. Combining with the binding slack function (i.e., $g_k(m^1, m^2; \eta) = 0$) there are $2R$ equations, and the supermarkups are exactly identified.

The imputation algorithm we use to recover the supermarkups tracks that provided in Appendix C.1. We use $\tilde{m}_R^1$ as the base supermarkup off of which we calculate supermarkup differences, so that $\widetilde{\Delta m_r^1} = \tilde{m}_r^1 - \tilde{m}_R^1$ and $\widetilde{\Delta m_r^2} = \tilde{m}_r^2 - \tilde{m}_R^1$. With size-specific supermarkups, there are twice as many supermarkups, so the computation time is approximately double that of the baseline model.

It is theoretically possible that a solution to the set of $2R$ equations provided above (i.e. an interior solution) does not exist. This could occur if searching for a solution under the assumption that i is the binding firm generates $g_j(m^1, m^2; \eta) < 0$ (i.e., that the IC constraint is violated for firm j), but searching for a solution using j as the binding firm generates $g_i(m^1, m^2; \eta) < 0$.[50] Restated loosely, one coalition firm might prefer higher supermarkups for smaller package sizes, and another coalition firm might prefer higher supermarkups for larger package sizes. This does not occur in our application because the relative shares of the coalition firms are similar for the smaller and larger package sizes.

Independent Regions

The baseline model connects the regions through the IC constraint in the leader's maximization problem. It is also possible to proceed under the assumption that each region is completely independent. In this alternative version of the model, deviation in any single region is followed by punishment in that region, rather than punishment in all regions. The leader has region-specific constrained maximization problems:

$$m_{rt}^*(\eta) = \arg\max_{m_{rt} \geq 0} \pi_{1rt}^{PL}(m_{rt}) \quad s.t. \quad g_{irt}(m_{rt}; \eta) \geq 0 \quad \forall i \in \mathbb{C} \tag{C.3}$$

where $g_{irt}(m_{rt}; \eta)$ is as defined in equation (6). Under the assumption of a constrained PLE, the solution can be found by searching for the supermarkup for which

$$g_{krt}(m_{rt}; \eta) = 0 \tag{C.4}$$

[50] Our code does not require the binding firm to be specified ex ante, although doing so speeds computation by reducing the number of numerical derivatives that must be calculated.

where k is the coalition firm with the least slack. As the numerical search involves one equation, one unknown, and a well-behaved loss function, a solution can be found quickly. In practice, one distinction with this version of the model is that the identity of the binding firm is region-specific—in some regions Miller or ABI have the binding IC constraint.

C.3 Numerical Evidence on Imputation

Under the maintained assumptions, it is possible to recover marginal costs from data on prices and quantities given knowledge of demand, the identity of coalition firms, the timing parameter, and whether IC constraints bind. We conduct a numerical experiment to confirm that our coding approach accomplishes this in practice. To do this, we consider 100 markets and, in each, take the following steps:

1. **Data Generation:** Randomly draw the number of firms using equal probabilities for $4, 5, \ldots, 10$ firms. Assume logit demand generated by indirect utility $u_{ij} = -p_j + \xi_j + \epsilon_{ij}$. Randomly draw each product's quality ξ_j independently from a uniform distribution with support over [1,2]. Draw each product's marginal cost ω_j independently from a uniform distribution with support over [0,1]. Assume that the first two firms are in the coalition and the third firm is in the fringe. Randomly assign all other firms to the coalition with probability 0.50 and to the fringe with probability 0.50.

2. **Equilibrium Computation:** Compute the solution to the constrained maximization problem of the leader (firm 1), for each of the timing parameters $\eta = 0.2, 0.3, \ldots, 0.8$, using the fmincon function in Matlab. This obtains the prices, quantities, and supermarkups that arise in the PLE, as well as the Lagrangian. Let the IC constraints bind if the Lagrangian exceeds 1e-3.

3. **Imputation Algorithm:** Given the PLE prices and quantities, the identity of coalition firms, the timing parameter, and whether the IC constraints bind, recover the marginal costs using the fminsearch function in Matlab. This involves considering a candidate supermarkup, applying Proposition 1 to recover the implied marginal costs and Bertrand prices, evaluating either the slack functions at the observed PLE prices (with binding IC constraints) or the leader's profit function at the observed PLE prices (otherwise), and iterating to convergence.

Given each set of data generated in step 1, the PLE that we recover in step 2 features a positive supermarkup. The question is whether the marginal costs in the generated data and those from step 3 (imputation) align. Given the 100 markets, seven timing parameters, and the random draw on the number of firms, we obtain a total of 4,725 marginal costs, corresponding to 6.75 firms per market, on average. We find that 93.81% of the imputed marginal costs are within 0.1% of the data, and 98.86% of the imputed marginal costs are within 1% of the data. Thus, the numerical experiment corroborates that our coding approach accurately recovers marginal costs.

C.4 Simulation

We simulate the baseline model using methods similar to those we use in calibration. For the merger simulations, we maintain the assumption pooled regions in the IC constraints, so methods track Appendix C.1. The same first order conditions (equations (C.1) and (C.2)) apply, albeit with a modified mapping of products to firms. We also use the same algorithm to recover the supermarkups. For the examination of multimarket contact, we compare the baseline calibration results to those obtained by simulating outcomes with independent regions. Thus, methods track the approach provided in Appendix C.2.

The main difference between calibration and simulation pertains to how the profit terms that appear in the first order conditions are calculated. With calibration, we apply Proposition 1 to recover marginal costs and Bertrand prices for each candidate supermarkup vector, and then compute deviation prices and obtain the profit terms. With simulation, we hold the marginal costs fixed at the level obtained from calibration. We compute Bertrand prices and profit, which is invariant to the supermarkup. Then, for each candidate supermarkup, we compute the implied price leadership and deviation prices given the marginal costs, and then obtain the profit terms.

D Results with Alternative Assumptions

D.1 Overview

In the baseline model, we assume that there is a single supermarkup that applies to all coalition products in a region, that IC constraints are pooled across markets, and that ABI is the pricing leader. The first assumption is a matter of expediency, as computation time increases in the number of supermarkups that must be imputed. The second assumption can be justified on the basis that price leadership is more profitable for the leader with pooled IC constraints. The third assumption is based on the qualitative evidence. In this section, we revisit these assumptions and demonstrate that our main results are robust to reasonable alternatives. In particular, we examine the following three alternative modeling assumptions:

1. *Size-specific supermarkups.* We assume that ABI sets two supermarkups in each region: one that applies to 6 and 12 packs, and another that applies to 24 packs. We make this distinction in part based on the observation that there is stronger import competition for the smaller package sizes (e.g., see Table 2).

2. *Independent regions.* We assume that ABI sets the supermarkup in each region subject to region-specific IC constraints. Thus, multimarket contact has no bearing on the prices that can be sustained in the PLE, and regions can be considered independently.

3. *MillerCoors is the leader.* We assume that MillerCoors solves the constrained maximization problem laid out in the text.

For the first two of these scenarios, we impute the supermarkups and marginal costs under the alternative models, using techniques that are described in Appendix C.2. As in the baseline model, imputation requires the timing parameter to be specified *ex ante*. However, different assumptions imply different marginal costs, so it is possible distinguish *ex post*. We use the same marginal cost function as the baseline model, and also apply the same identifying assumption, namely that $\beta_1 = 0$. We examine Bertrand competition, unconstrained PLE, and (initially) constrained PLE with $\eta = 0.25$ and $\eta = 0.30$. Based on these results, we project the timing parameter that would bring the estimate of β_1 closest to zero, and obtain an additional set of results for that timing parameter.

For the third scenario, we hold fixed the marginal costs that we obtain in the baseline model and simulate a new equilibrium under the counterfactual assumption that MillerCoors is the leader. The simulation methodologies are described in Appendix C.2.

D.2 Results

The calibration results generated with the alternative models are summarized in Appendix Tables G.1 and G.2. They are broadly consistent with the results from the baseline model (Table 3). With each of the alternative models, Bertrand and unconstrained PLE are rejected. The timing parameters that best satisfy the identifying assumption of $\beta_1 = 0$ are 0.25 and 0.27, respectively. The average supermarkups also are similar to those of the baseline model, so our main results are robust to some of the modeling specifics.

We now discuss the alternative models in greater detail. First, consider the model with size-specific supermarkups. The results of Table G.1 indicate that supermarkups are slightly higher for 24 packs than for 6 and 12 packs, a finding that is consistent with Modelo and Heineken have a smaller presence for the largest package sizes. However, the differences are not great. For example, in fiscal year 2007, we find an average supermarkup of 1.12 for the smaller package sizes and 1.23 for the larger package sizes. This suggests that the baseline model might be a good approximation even if the richer (and more computationally demanding) model with is correct.

Appendix Figure G.1 provides some scatter plots to explore this further. Panel A shows that, in the alternative model, there is a high degree of correlation between the region-year specific supermarkups for the smaller and larger package sizes. Thus, regions that have high supermarkups for one size group also have high supermarkups for the other group. Next, Panel B shows that the average of the region-year supermarkups in the alternative model (vertical axis) is highly correlated with the region-year supermarkups from the baseline model (horizontal axis). Together, Panels A and B indicate that the alternative model is similar to the baseline model in terms of the supermarkups that they imply. In principle, the models could nonetheless be distinguished on the basis that they produce (somewhat) different marginal costs. However, Panel C plots the marginal costs—there is enough similarity that distinguishing the models would be difficult in practice.

We now turn to the alternative model with independent regions, such that IC constraints are not pooled and multi-market contact does not matter for the results. As the IC constraints in the alternative model are (on average) the same as the IC constraints in the baseline model, it is perhaps not surprising that the supermarkups are also similar on average. More interesting is in how supermarkups are set across regions. With the baseline model, dispersion across regions satisfies first order conditions reflecting how the region-specific supermarkups affects (i) the leader's profit and (ii) the binding firm's incentive to deviate. With the alternative model, the region-supermarkups reflect the maximum supermarkup that can be sustained given the binding firm's IC constraint.

We compare the results for the two models in Appendix Figure G.2, using a series of scatter plots. Panels A and B show the region-specific supermarkups in fiscal years 2007 and 2010, respectively. In the earlier fiscal year, there are noticeable differences between the models, but these mostly disappear in the later fiscal year. We believe this change reflects the greater symmetry among the coalition firms that exists after the Miller/Coors merger. If the interests of the leader and the binding firm tend to align, then there appears to be less scope for multi-market contact to affect results. Panel C shows the marginal costs that obtain with the two models. There are some minor differences for the less costly products (i.e., the domestic products) but also a high degree of correlation overall. Again, this variation could in principle distinguish the models, but in practice it appears insufficient to generate reliable results. Thus, in our view, the obtained marginal costs do not provide a good statistical basis for preferring one model over the other.

Finally, in Appendix Figure G.3, we plot the supermarkups from the baseline model, in which ABI is the pricing leader, against those obtained from a counterfactual simulation in which we let MillerCoors be the leader. The fiscal year for the comparison is 2010. As shown, the supermarkups do not depend much on the identity of the leader, as the dots are clustered around the 45° line. This is another implication of the overall symmetry between ABI and MillerCoors in the latter part of our sample.

E The Demand System

Here we sketch the Miller and Weinberg (2017) random coefficients nested logit (RCNL) model of demand. Suppose we observe $r = 1, \ldots, R$ regions over $t = 1, \ldots, T$ time periods. Each consumer i purchases one of the observed products ($j = 1, \ldots, J_{rt}$) or selects the outside option ($j = 0$). The conditional indirect utility that consumer i receives from the inside good j in region r and period t is

$$u_{ijrt} \;=\; x_j \beta_i^* - \alpha_i^* p_{jrt} + \sigma_j^D + \tau_t^D + \xi_{jrt} + \bar{\epsilon}_{ijrt} \tag{E.1}$$

where x_j is a vector of observable product characteristics, p_{jrt} is the retail price, σ_j^D is the mean valuation of unobserved product characteristics, τ_t^D is the period-specific mean valuation of unobservables that is common among all inside goods, ξ_{jrt} is a region-period deviation from these means, and $\bar{\epsilon}_{ijrt}$ is a mean-zero stochastic term.

The observable product characteristics include a constant (which equals one for the inside goods), calories, package size, and an indicator for whether the product is imported. The consumer-specific coefficients are $[\alpha_i^*, \beta_i^*]' = [\alpha, \beta]' + \Pi D_i$ where D_i is consumer income. Define two groups, $g = 0, 1$, such that group 1 includes the inside goods and group 0 is the outside good. Then the stochastic term is decomposed according to

$$\bar{\epsilon}_{ijrt} = \zeta_{igrt} + (1 - \rho)\epsilon_{ijrt} \tag{E.2}$$

where ϵ_{ijrt} is i.i.d extreme value, ζ_{igrt} has the unique distribution such that $\bar{\epsilon}_{ijrt}$ is extreme value, and ρ is a nesting parameter ($0 \leq \rho < 1$). Larger values of ρ correspond to less substitution between the inside and outside goods. The quantity sold of good j in region r and period t is

$$q_{jrt} = \frac{1}{N_{rt}} \sum_{i=1}^{N_{rt}} \frac{\exp((\delta_{jrt} + \mu_{ijrt})/(1-\rho))}{\exp(I_{igrt}/(1-\rho))} \frac{\exp(I_{igrt})}{\exp(I_{irt})} M_r \tag{E.3}$$

where I_{igrt} and I_{irt} are the McFadden (1978) inclusive values, M_r is the market size of the region, $\delta_{jrt} = x_j\beta + \alpha p_{jrt} + \sigma_j^D + \tau_t^D + \xi_{jrt}$, and $\mu_{ijrt} = [p_{jrt}, x_j]' * \Pi D_i$. The normalization on the mean indirect utility of the outside good yields $I_{i0rt} = 0$. The inclusive value of the inside goods is $I_{i1rt} = (1-\rho)\log\left(\sum_{j=1}^{J_{rt}} \exp((\delta_{jrt} + \mu_{ijrt})/(1-\rho))\right)$ and the inclusive value of all goods is $I_{irt} = \log(1 + \exp(I_{i1rt}))$. We assume market sizes 50% greater than the maximum observed unit sales within each region. Expressions for the price derivatives of demand are supplied in Grigolon and Verboven (2014).

The parameters are estimated with GMM. The general approach follows the standard nested fixed-point algorithm (Berry et al. (1995)), albeit with a modification to ensure a contraction mapping in the presence of the nested logit structure (Grigolon and Verboven (2014)). As demand estimation is not the primary focus of this paper, we refer readers to Miller and Weinberg (2017) for the details of implementation, a discussion of the identifying assumptions, specification tests, and a number of robustness analyses.

F Adding a Retail Sector

The baseline model assumes that brewers set prices to consumers. However, it is identical to a model that incorporates a constant-markup retail sector (i.e., one that uses "cost-plus" pricing). The reason is that retail markups and brewer marginal costs enter the model in the same way. Although cost-plus pricing does not maximize retail profit, it may provide a reasonable rule-of-thumb policy, and recent research provides some support for it using scanner data similar to the IRI data that we use. In particular, DellaVigna and Gentzkow (2019) shows that retail prices often do not respond to local demand shocks, and Butters et al. (2020) documents that retail prices change one-for-one with local cost shocks (generated by excise taxes). This combination would arise from cost-plus pricing.

With the constant-markup retail sector, the profit function of the brewer i takes the form:

$$\pi_i(p^w) = \sum_{j \in \mathbb{J}_i} (p_j^w - mc_j) q_j(p^r(p^w)) \tag{F.1}$$

where p^w and p^r are the brewer and retail prices, respectively. (We have suppressed subscripts for the region and period for simplicity). Retail prices are set according to $p^r = p^w + \mu^r$ where μ^r is a vector of product-specific markups. Define $\mu^w \equiv p^w - mc$ as the brewer markup. Thus, we have $p^r = mc + \mu^w + \mu^r$. Now, as marginal costs are constant, profit can be re-expressed as a function of brewer markups:

$$\begin{aligned} \pi_i(\mu^w) &= \sum_{j \in \mathbb{J}_i} \mu_j^w q_j(mc + \mu^w + \mu^r) & \text{(F.2)} \\ &= \sum_{j \in \mathbb{J}_i} \mu^w q_j(\mu^w + \phi) & \text{(F.3)} \end{aligned}$$

where $\phi = mc + \mu^r$ is the combined retail markup and brewer cost. From the perspective of the brewer, ϕ matters for profit but the allocation of ϕ between retail markup and brewer cost does not. The static first order conditions, the slack functions, and the leader's constrained maximization problem all work with this profit function in various ways, and can be recast in terms of brewer markups, yielding the same isomorphism. Thus, the baseline price leadership model is equivalent to an alternative with cost-plus retail pricing.

Miller and Weinberg (2017, Appendix E) also consider a profit-maximizing retailer in a model that is similar in some respects to the one used here, and find that supply-side inferences about brewer markups are robust. In that setting, retail markups are not constant, but they are close enough to constant that they affect brewer incentive in nearly the same way as a marginal cost shifter. Incorporating a profit-maximizing retailer into the price leadership model would greatly complicate imputation, enough so that we believe it would be prohibitive given the algorithms we employ.

G Additional Tables and Figures

Table G.1: Model Selection with Size-Specific Supermarkups

Panel A: OLS Regression Results					
	Bertrand	$\eta = 0.25$	$\eta = 0.26$	$\eta = 0.30$	Unconstrained
ABI×Post-Merger	0.657	0.009	-0.031	-0.210	0.504
	(0.094)	(0.118)	(0.121)	(0.122)	(0.147)
Miller×Post-Merger	0.189	-0.480	-0.523	-0.711	-0.037
	(0.074)	(0.012)	(0.099)	(0.103)	(0.133)
Coors×Post-Merger	-0.067	-0.756	-0.798	-0.986	-0.278
	(0.095)	(0.110)	(0.110)	(0.117)	(0.169)
Distance	0.22	0.231	0.232	0.237	0.331
	(0.066)	(0.066)	(0.067)	(0.068)	(0.120)

Panel B: Other Statistics					
	Bertrand	$\eta = 0.25$	$\eta = 0.26$	$\eta = 0.30$	$\eta = 1.00$
$\overline{m}_{2006}^{6,12}$	0	1.02	1.07	1.27	3.99
$\overline{m}_{2006}^{24}$	0	1.13	1.18	1.43	5.08
$\overline{m}_{2007}^{6,12}$	0	1.07	1.12	1.33	4.11
$\overline{m}_{2007}^{24}$	0	1.18	1.24	1.49	5.18
$\overline{m}_{2010}^{6,12}$	0	1.64	1.73	2.10	4.24
$\overline{m}_{2010}^{24}$	0	1.71	1.80	2.22	4.94
$\overline{m}_{2011}^{6,12}$	0	1.72	1.81	2.20	4.54
$\overline{m}_{2011}^{24}$	0	1.75	1.84	2.27	4.97
Marginal Costs < 0	0.001	0.005	0.006	0.010	0.163

Notes: Panel A summarizes results from OLS estimation of the marginal cost function. The marginal costs are imputed using a model in which ABI selects one supermarkup for 6 and 12 packs, and another for 24 packs. The IC constraints are pooled across regions. The intermediate timing parameters are generated under the assumption that the (pooled) IC constraint binds. Regressors include indicators for ABI brands, Miller brands, and Coors brands, in the fiscal years 2010 and 2011 (corresponding to β_1, β_2, and β_3), distance from the brewery to the region (β_4), as well as product, region, and time fixed effects. Standard errors are clustered at the region level and shown in parentheses. Panel B provides the average supermarkups by fiscal year and by pack size, as well as the proportion of marginal cost values which are negative.

Table G.2: Model Selection with Independent Regions

	Panel A: OLS Regression Results				
	Bertrand	$\eta = 0.25$	$\eta = 0.27$	$\eta = 0.30$	Unconstrained
ABI×Post-Merger	0.657	0.087	0.012	-0.114	0.607
	(0.094)	(0.096)	(0.097)	(0.099)	(0.147)
Miller×Post-Merger	0.189	-0.429	-0.508	-0.64	-0.017
	(0.074)	(0.068)	(0.069)	(0.070)	(0.137)
Coors×Post-Merger	-0.067	-0.698	-0.778	-0.911	-0.242
	(0.095)	(0.097)	(0.098)	(0.101)	(0.173)
Distance	0.22	0.227	0.227	0.228	0.34
	(0.066)	(0.067)	(0.067)	(0.067)	(0.120)
	Panel B: Other Statistics				
	Bertrand	$\eta = 0.25$	$\eta = 0.27$	$\eta = 0.30$	$\eta = 1.00$
$\overline{m}_{2006}$	0	1.11	1.22	1.39	4.51
$\overline{m}_{2007}$	0	1.14	1.25	1.43	4.65
$\overline{m}_{2010}$	0	1.64	1.82	2.11	4.63
$\overline{m}_{2011}$	0	1.7	1.89	2.19	4.79
Marginal Costs < 0	0.001	0.005	0.006	0.009	0.17

Notes: Panel A summarizes results from OLS estimation of the marginal cost function. The IC constraints are *not* pooled across regions. Regressors include indicators for ABI brands, Miller brands, and Coors brands, in the fiscal years 2010 and 2011 (corresponding to β_1, β_2, and β_3), distance from the brewery to the region (β_4), as well as product, region, and time fixed effects. The intermediate timing parameters are generated under the assumption that the IC constraints bind. Standard errors are clustered at the region level and shown in parentheses. Panel B provides the average supermarkups by fiscal year, as well as the proportion of marginal cost values which are negative.

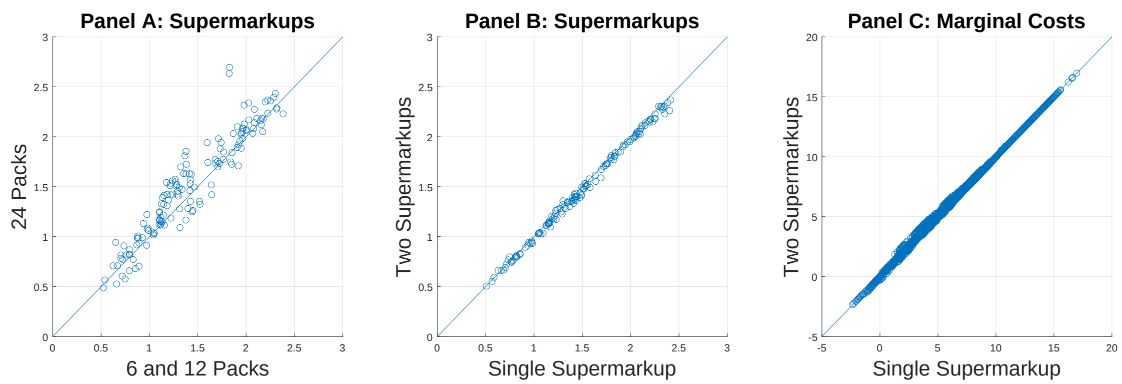

Figure G.1: Size-Specific Supermarkups and the Baseline Model

Notes: Panel A plots the supermarkups imputed using the alternative assumption that, in each region, the leader sets one supermarkup for 6 and 12 packs, and one supermarkup for 24 packs. Each dot is a region-fiscal year combination and shows the values of the two supermarkups. Panel B is compares the supermarkups imputed using this alternative assumption to those imputed using the baseline assumption of a single supermarkup. Each dot is a region-fiscal year combination. The vertical axis is the mean supermarkup in the two-supermarkup model. Panel C plots the marginal costs imputed using the alternative and baseline assumptions. Each dot is a product-region-period combination. Each panel has a 45° line to assist with interpretation.

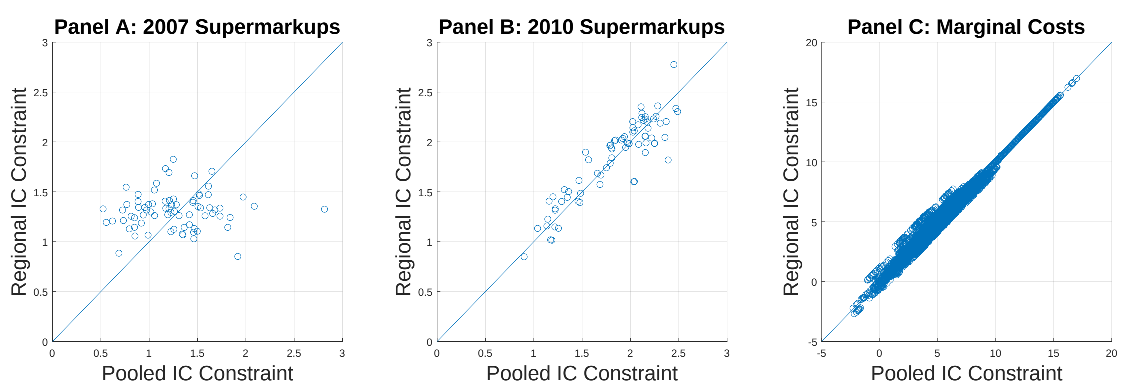

Figure G.2: Regional IC Constraints vs. the Baseline Model

Notes: Panel A plots the supermarkups imputed using the alternative assumption of regional IC constraints for the fiscal year 2007, against those imputed using the baseline assumption of pooled IC constraints. Each dot is a region. Panel B is identical except that it shows the fiscal year 2010. Panel C plots the marginal costs imputed using the alternative and baseline assumptions. Each dot is a product-region-period combination. Each panel has a 45° line to assist with interpretation.

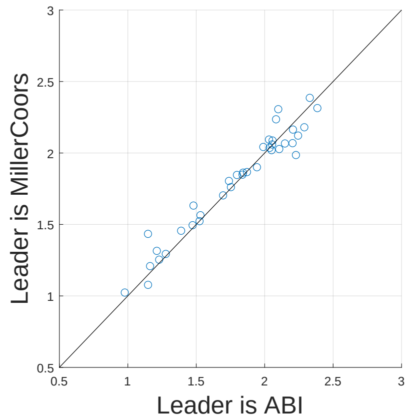

Figure G.3: Supermarkups with ABI and MillerCoors as Leader

Notes: The figure plots the supermarkups obtained from a counterfactual simulation in which MillerCoors is the leader (vertical axis) against the supermarkups obtained from the baseline model in which ABI is the leader (horizontal axis). Each dot is a region in fiscal year 2010. The figure includes a 45° line.

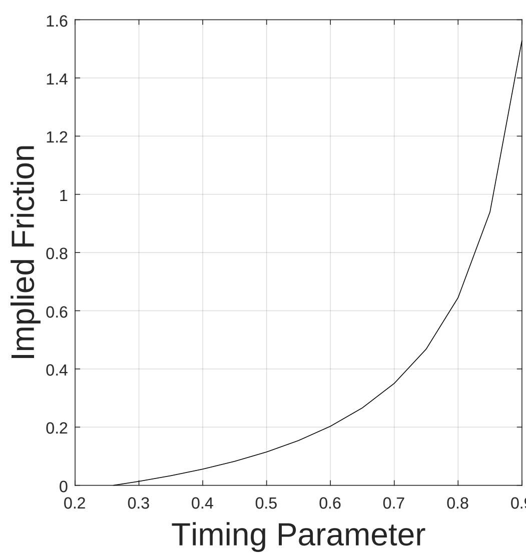

Figure G.4: Slack Functions in the Numerical Illustration

Notes: The figure illustrates joint identification of the timing parameter and a coordination friction, under the augmented IC constraints discussed in Section 5. The implied friction is expressed as a proportion of MillerCoors' price leadership profit. The exercise is conducted for fiscal year 2010. With $\eta = 0.45$, the magnitude of the friction is equivalent to 8.25% of MillerCoors' profit in the PLE.