Understanding the Price Effects of the MillerCoors Joint Venture*

Nathan H. Miller[†]
Georgetown University

Matthew C. Weinberg[‡]
Drexel University

February 24, 2017

Abstract

We document abrupt increases in retail beer prices just after the consummation of the MillerCoors joint venture, both for MillerCoors and its major competitor, Anheuser-Busch. Within the context of a differentiated–products pricing model, we test and reject the hypothesis that the price increases can be explained by movement from one Nash-Bertrand equilibrium to another. Counterfactual simulations imply that prices after the joint venture are 6%–8% higher than they would have been with Nash–Bertrand competition, and that markups are 17%–18% higher. We relate the results to documentary evidence that the joint venture may have facilitated price coordination.

Keywords: Market power; mergers; unilateral effects; coordinated effects; antitrust policy; merger enforcement; brewing industry
JEL classification: K21; L13; L41; L66

---

*We thank Jonathan Baker, Dan Hosken, Robert Porter, Ted Rosenbaum, Chuck Romeo, Gloria Sheu, and Jonathan Williams for detailed comments, along with seminar participants at various economics departments and business schools. We have also benefited from conversations with John Asker, Orley Ashenfelter, Michael Baye, Hank Farber, Allan Collard-Wexler, Michael Grubb, Ali Hortaçsu, J.F. Houde, Aviv Nevo, and Mike Waldman, among others. Conor Ryan and Andrew Gellert provided careful research assistance. All estimates and analyses in this paper based on SymphonyIRI Group, Inc., data are by the authors and not by SymphonyIRI Group, Inc.

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

[‡]Drexel University, Gerri C. LeBow Hall, 3220 Market Street, Philadelphia PA 19104. Email: mcw325@drexel.edu.

1 Introduction

Economic theory indicates that repeated interaction within oligopolies can support collusive equilibria if there are few enough firms (e.g., Friedman (1971), Abreu (1988)). Accordingly, the Merger Guidelines of the United States Department of Justice (DOJ) and Federal Trade Commission (FTC) emphasize that mergers in concentrated markets can facilitate coordination and nearly 60% of merger complaints filed by the DOJ and FTC over 1990–2014 allege coordinated effects (Gilbert and Greene (2015)). Most empirical research, by contrast, focuses on the unilateral effects of mergers that arise from the internalization of diverted sales between merging firms selling differentiated brands (e.g., Berry and Pakes (1993); Hausman, Leonard and Zona (1994); Werden and Froeb (1994); Nevo (2000a)). In these models, post-merger coordination is not considered. Instead, firms are assumed to play one-shot Nash-Bertrand pricing equilibria both before and after the merger.

We study the economic effects of MillerCoors, a joint venture of SABMiller PLC and Molson Coors Brewing that combined the operations of these brewers in the United States. The joint venture underwent antitrust review as a merger between the second and third largest firms in the U.S. brewing industry. It was approved June 5, 2008 by the DOJ on the basis that merger-specific cost reductions would likely outweigh any anticompetitive effects. Normal course documents of Anheuser-Busch InBev (ABI)—the closest competitor of MillerCoors—released publicly in later antitrust litigation describe the goals of the company's pricing practices as "yielding the highest level of followership in the short-term" and "improving competitor conduct over the long-term." While business documents must be interpreted carefully, this language raises the question of whether the Miller/Coors merger facilitated coordination in the brewing industry.

We start with a descriptive analysis of retail price data that span 39 geographic regions over years 2001-2011. Inflation-adjusted prices are stable around a small downward trend over the seven years preceding the merger. The prices of MillerCoors and ABI then increase abruptly in the fall of 2008, just after the Miller/Coors joint venture, and this increase persists through the end of the sample period. We estimate the magnitude of the price increase to be roughly six percent. Regions in which MillerCoors and ABI have greater pre-merger market shares experienced larger price increases, whereas regions expected to have greater merger-specific cost reductions experienced smaller price increases. These opposing forces nearly offset each other in the average market, leaving unexplained a four percent increase in MillerCoors and ABI prices that is common across all regions.

To explore the possibility of coordinated effects, we estimate a structural model of demand and supply that allows for post-merger departures from Nash–Bertrand competition. The supply side of the model incorporates a parameter that determines the extent to which MillerCoors and ABI internalize their pricing externality during the post-merger periods. The other competitors in the model—Modelo and Heineken—are assumed to compete *à la* Nash–Bertrand both against each other and against ABI/MillerCoors. This demarcation between the domestic and import competitors is supported both in the data and in qualitative evidence that we summarize later. The demand side of the model is standard. We use a discrete choice random utility model that allows for the estimation of reasonable consumer substitution patterns with aggregated data on market shares and prices. Similar models have been applied to the beer industry before (e.g., Hellerstein (2008); Goldberg and Hellerstein (2013); Asker (2016); Romeo (2016); Sweeting and Tao (2016)).

The model rejects Nash–Bertrand competition if the post-merger prices of ABI exceed what can be explained by unilateral effects. The focus on ABI allows us to flexibly capture merger-specific cost reductions for MillerCoors. Because even large increases in ABI's prices could be rationalized by some unobserved shock, restrictions must be placed on the structural error terms. The key identifying assumption is that changes in ABI's unobserved demand and costs, before versus after the merger, are not systematically different from changes in the unobserved demand and costs of Modelo and Heineken. We interpret the merger itself as a plausibly exogenous shifter of the competitive environment (Berry and Haile (2014)) and use this assumption to form moment conditions. The strategy benefits from the presence of competitors outside of the coordinating group, which allows us to control for unobserved demand and cost changes that are common across firms.

The results are consistent with the Miller/Coors merger having coordinated effects. The governing supply-side parameter is statistically different from zero and robust across a number of modeling choices. The model thus rejects Nash–Bertrand competition in the post-merger periods. Strictly interpreted, the point estimate on our preferred specification implies that ABI and MillerCoors internalize 26% of their pricing externalities after the merger. Using counterfactual simulations, we determine that the observed post-merger prices of these firms are six to eight percent higher than they would have been under Nash–Bertrand competition and markups are 17%–18% higher. We quantify the loss of consumer surplus and show that it is attributable primarily to softer price competition; the downward pricing pressure from merger-specific cost reductions and the upward pricing pressure from unilateral effects are roughly the same magnitude. Nonetheless, the cost reductions are large enough that the merger increases total surplus.

Our analysis is subject to a number of caveats and limitations. We highlight three here. First, the joint venture occurred during a turbulent time period in both the overall economy and the brewing industry. We control for changes in the market environment to the extent possible, but acknowledge that events other than Miller/Coors merger could have contributed to the observed price increases. Second, we could misattribute price increases resulting from unilateral effects to coordination if the model understates the degree to which ABI and MillerCoors prices are strategic complements in Nash–Bertrand equilibrium. This could happen if, for instance, brewers have imperfect information about each others' costs and play a dynamic signaling game (Sweeting and Tao (2016)).[1] Finally, we do not provide a formal theory that explains why the Miller/Coors joint venture may have enabled coordination. Instead, we provide a qualitative discussion about the brewing industry and the characteristics enumerated in the Merger Guidelines as contributing to coordinated effects.

This paper relates to several areas of research. The methodology is closest to that of Ciliberto and Williams (2014), who model the airline industry as a differentiated-products pricing game. Departures from Nash–Bertrand are assumed to be proportional to the multi-route contact between carriers. Inference depends on whether prices are higher on routes that feature more multi-route contact relative to what would arise in Nash–Bertrand equilibrium. Our identification strategy is also similar to that of Porter (1983) and Igami (2015), who focus on regime shifts in markets with homogeneous products. The former article examines the railroad industry and estimates reversions between competitive and collusive regimes. Under the assumption of undifferentiated Bertrand equilibrium in competitive regimes, the estimates obtained suggest roughly Cournot levels of output in the collusive regimes. Igami examines a cartel in the coffee bean industry. Inferences are made based on the magnitude of the price decreases that occur after the cartel collapses, under the assumption that post-cartel competition is Nash–Cournot.

A growing number of articles examine *ex post* effects of mergers (for a survey, see Ashenfelter, Hosken and Weinberg (2014)). Most commonly these "merger retrospectives" employ program evaluation techniques to estimate price changes.[2] A handful of studies compare these changes to the predictions obtained under the assumption of Nash–Bertrand competition before and after the merger (e.g., Peters (2006); Weinberg and Hosken (2013);

[1] In the Sweeting and Tao (2016) model, firms set higher prices than with perfect information Nash–Bertrand competition in order to signal high costs. This incentive is weak when there are many firms but can be strong with few firms. Mergers amplify the degree to which prices exceed Nash–Bertrand levels.

[2] One such article provides evidence on how prices changed across different geographic markets after the Miller/CoorsMiller–Coors merger (e.g., Ashenfelter, Hosken and Weinberg (2015)). The article explains variation in price changes across regions using how the merger would increase local market concentration and reduce shipping distances. It finds a negative relation between prices and shipping distances and a positive relation between prices and concentration. The same empirical patterns exist in our data.

Houde (2012); Bjornerstedt and Verboven (2015)). One recent working paper examines a merger in the ready-to-eat cereal industry and seeks to identify departures from Nash–Bertrand competition (Michel (2016)). The analysis of whether mergers lead to changes in how firms compete is still novel in this literature and could help account for discrepancies between the predictions of merger simulation and how prices actually change after mergers.

Finally, our article relates to research that measures market power using price and quantity data and at most incomplete data on costs. Because our test for post-merger coordination is based on whether changes in ABI prices can be explained with unilateral effects, it is crucial that any strategic complementarity of MillerCoors and ABI prices be captured in a reasonable manner. We therefore follow Nevo (2001) and implement the test after estimating a random utility model of demand that allows for flexible consumer substitution patterns. Further, as in Bresnahan (1987), and Nevo (2001), we assess the plausibility of different models of competition by examining their implied unit costs of production. If we impose Nash–Bertrand competition after the merger, taking into account the unilateral effects, a 13 percent increase in ABI costs is needed to rationalize the data. Publicly available company documents do not support such a cost increase.

The rest of the article is organized as follows. Section 2 provides background information on the U.S. brewing industry and the datasets used in the analysis. Section 3 examines variation in changes in retail prices before and after the merger and summarizes a body of qualitative evidence regarding coordination between MillerCoors and ABI. Section 4 develops the demand model and discusses the results of estimation. Section 5 develops the supply model and discusses the supply-side estimation results. Section 6 quantifies the economic importance of deviations from Nash–Bertrand competition using counterfactual simulations. Section 7 relates our empirical approach to the conduct parameter literature. Section 8 concludes with a discussion of distinctive features of the U.S. brewing industry that may have led the merger to soften competition beyond what can be explained with unilateral effects. Robustness analysis and extensions are available in the appendix.

2 Industry background

2.1 Market structure

As do most firms in branded consumer product industries, brewers compete in prices, new product introductions, advertising, and periodic sales. The beer industry differs from typical retail consumer product industries in its vertical structure because of state laws regulating

Table I: Revenue Shares and HHI

Year	ABI	MillerCoors	Miller	Coors	Modelo	Heineken	Total	HHI
2001	0.37	.	0.20	0.12	0.08	0.04	0.81	2,043
2003	0.39	.	0.19	0.11	0.08	0.05	0.82	2,092
2005	0.36	.	0.19	0.11	0.09	0.05	0.79	1,907
2007	0.35	.	0.18	0.11	0.10	0.06	0.80	1,853
2009	0.37	0.29	.	.	0.09	0.05	0.80	2,350
2011	0.35	0.28	.	.	0.09	0.07	0.79	2,162

Notes: This table provides revenue shares and the HHI over 2001–2011. Firm-specific revenue shares are provided for ABI, Miller, Coors, Modelo, and Heineken. The total across these firms is also provided. The HHI is scaled from 0 to $10,000$. The revenue shares incorporate changes in brand ownership during the sample period, including the merger of Anheuser-Busch (AB) and InBev to form ABI, which closed in November 2008, and Heineken's acquisition of the FEMSA brands in April 2010. All statistics are based on supermarket sales recorded in IRI scanner data.

the sales and distribution of alcohol. 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.[3] 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 I shows revenue-based market shares at two-year intervals over 2001–2011, based on retail scanner data that we describe later in this section. The brands of five brewers—ABI, SABMiller, Molson Coors, Grupo Modelo, and Heineken—account for approximately 80% of total retail revenue and there is no obvious downward trend in this revenue share, despite the recent growth of microbreweries. ABI accounts for about 35% of retail revenue and MillerCoors accounts for around 30%. Modelo and Heineken, both importers, together account for about 15% of revenues. The national Herfindahl–Hirschman Indexes (HHIs) are in the range that characterizes "moderately concentrated" markets in the DOJ–FTC Merger Guidelines. Many regions exhibit greater concentration due to distinct supply and demand conditions: 23 of the 39 regions in our sample have HHIs above 2,500 in 2011, which is in the range that characterizes highly concentrated markets in the Merger Guidelines.

Consolidation in the industry has continued since the Miller/Coors merger. ABI acquired Modelo in 2013, after our sample period. The DOJ obtained a settlement in which the rights to produce and distribute Modelo brands in the United States were divested to Con-

[3]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).

stellation, a large distributor and producer of wine and spirits. Subsequently, ABI acquired SABMiller itself in a deal worth $106 billion that closed in October 2016. The divestiture package requires the merged entity to divest the SABMiller stake in MillerCoors.

2.2 Data sources

Our primary data source is retail scanner data from the IRI Academic Database (Bronnenberg, Kruger and Mela (2008)). The data include revenue and unit sales by UPC code, week, and store for a sample of supermarkets over 2001–2011. We restrict the regression samples to 39 distinct geographic regions and 13 flagship brands. These brands 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. The most popular brands that we omit are either regional brands (e.g., Yuengling) or in the "subpremium" category and sell at much lower price points.

Beer is sold in different package sizes and hereafter we refer to brand×size combinations as products. We focus on six packs, 12 packs, 24 packs, and 30 packs. Thus, "Bud Light 12 Pack" is one product in the sample. We combine 24 packs and 30 packs in the construction of our products because whether 24 packs or 30 packs are sold tends to depend on region-specific historical considerations. We exclude 18-packs and promotional package sizes, which are much less popular. Following standard practice, we measure market shares based on 144-ounce equivalent units, the size of a 12 pack. This means, for example, that the sale of a six pack is downweighted by 50% in the construction of market shares. Prices are then defined as the ratio of revenue to equivalent unit sales. Typically the larger package sizes are less expensive on an equivalent unit basis. In total, 12 packs produce the greatest number of unit sales and 24 packs account for the greatest sales volume.

For computational reasons, we aggregate the data from the store–week level to the region–month and region–quarter levels. A potential concern with our static approach to estimating demand is that sales and consumer stockpiling could cause a bias that understates unilateral incentives to raise prices (Hendel and Nevo (2006)). While aggregation over time reduces this bias only in special cases, we vary the periodicity of the sample in this way to provide some assurance that dynamic considerations do not drive the results. The identification strategy does not require week-to-week variation in the data, so this aggregation may even be helpful insofar as it reduces random measurement error. We revisit our descriptive regressions in Appendix B using store-level observations and show that the main empirical patterns are not created by changes in the store-level composition of the IRI data. There are 167,695 observations at the product–region–month–year level, spanning 2001–2011.

We use household demographics from the Public Use Microdata Sample (PUMS) of the American Community Survey to help estimate demand. 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. We merge the PUMS data into the IRI scanner data by matching on the counties that compose the IRI regions and the PUMS areas. In estimation, we take 500 draws on households per region–year and obtain household income as total income divided by the number of household members. The mean income is $38 thousand dollars. When using the PUMS, we necessarily focus on the 2005–2011 period. We also discard data from the first year following the Miller/Coors merger to allow for the realization of cost reductions. There are 94,656 qualifying observations at the product–region–month–year level and 31,784 observations at the product–region–quarter–year level.

We obtain the driving miles between each IRI region and the nearest brewery for each product in our sample using Google Maps. For imported brands, we define the miles traveled based on the nearest port into which the beer is shipped.[4] We construct a notion of distance based on the interaction of driving miles and diesel fuel prices, which we obtain from the U.S. Energy Information Administration of the U.S. 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 intertemporally, based on fluctuations in fuel costs. It also helps us estimate the distributional cost savings of the Miller/Coors merger. All prices and incomes are deflated using the Consumer Price Index and are reported in 2010 dollars.

3 Retail prices

3.1 Time series variation

Figure 1 plots average log retail prices over 2001–2011 for each firm's best selling 12 pack: Bud Light, Miller Lite, Coors Light, Corona Extra, and Heineken. The vertical line at June 2008 signifies the consummation of the Miller/Coors merger. Horizontal ticks are placed at October because brewers typically adjust their prices in early autumn. Retail prices trend

[4] 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 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.

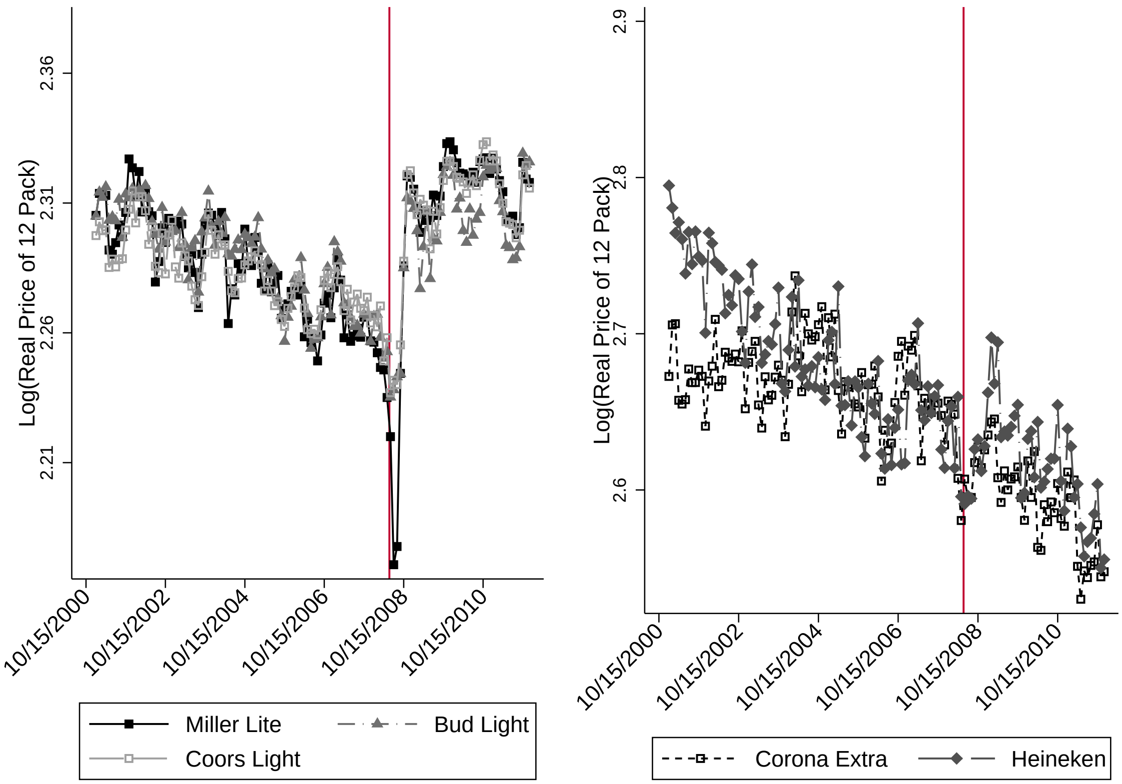

Figure 1: Average Retail Prices of Flagship Brand 12 Packs

Notes: This figure plots the average prices 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.

downward before the merger for all five products, a period spanning more than seven years. After the merger, the prices of Bud Light, Miller Lite, and Coors Light increase by about 8% and there is no obvious continuation of the downward trend. The prices of Corona Extra and Heineken do not exhibit any persistent increase and instead continue along a downward trend. The price gap between the cheaper domestic beers and the more expensive imports shrinks over time in the post-merger periods.

The most theoretically interesting aspects of Figure 1 are that (i) the price of Bud Light increases by roughly the same amount as the prices of Miller Lite and Coors Light and (ii) Modelo and Heineken prices do not increase, at least not persistently. Post-merger coordination between ABI and MillerCoors is one possible explanation. Alternatively, the data could be explained solely by unilateral effects, under a particular set of demand elasticities that produces strong strategic complementarity among the prices of domestic beers and weak strategic complementarity between the prices of domestic and imported beers. Specific institutional practices could also be important. As one example, retailers could set equal prices for Bud Light, Miller Lite, and Coors Light, regardless of external circumstances, due to beliefs about the market or pressure from the brewers. Changing macroeconomic conditions are also relevant because the merger coincides with the onset of the Great Recession. Income losses could decrease the demand elasticities of domestic beer for a variety of reasons, including down-market substitution.[5]

We use difference-in-differences regressions to quantify the changes and expand inference to the other flagship brands. The following regression equation specifies the log retail price of product j in region r in period t according to:

$$\begin{aligned} \log p^R_{jrt} &= \beta_1 \mathbb{1}\{\text{MillerCoors}\}_{jt} \times \mathbb{1}\{\text{Post-Merger}\}_t \\ &+ \beta_2 \mathbb{1}\{\text{ABI}\}_{jt} \times \mathbb{1}\{\text{Post-Merger}\}_t \\ &+ \beta_3 \mathbb{1}\{\text{Post-Merger}\}_t + \phi_{jr} + \tau_t + \epsilon_{jrt} \end{aligned} \quad (1)$$

which includes indicator variables for (i) MillerCoors brands in the post-merger periods, (ii) ABI brands in the post-merger periods, and (iii) all products in the post-merger periods. We absorb cross-sectional variation with product $\times$ region fixed effects (the ϕ_{jr} parameters). We use a linear time trend (τ_t) to account for the secular downward trend in prices. The alternative of time fixed effects produces nearly identical results, though the fixed effects prevent the inclusion of the post-merger indicator $\mathbb{1}\{\text{Post-Merger}\}_t$. In some specifications, we expand on equation (14) by allowing the linear trend to vary freely across products and by adding quarterly employment levels and average weekly earnings from the Quarterly Census of Employment and Wages to help capture local economic conditions.

Table II presents the regression results. The sample in column (i) includes 12 packs of Bud Light, Coors Light, Miller Lite, Heineken, and Corona and so corresponds precisely to Figure 1. The coefficients indicate that MillerCoors and ABI prices increase by an average of 10.3 percent (because $\exp(0.098) - 1 = 0.103$) and 9.1 percent, respectively, relative to imported brands; the absolute price increases are, respectively, 6.8 and 5.7 percent. The difference between the MillerCoors and ABI coefficients is not statistically significant. Column

[5] The observed price patterns probably are not due exclusively to down-market substitution, however, because the sales of the domestic brands decrease in both absolute and relative terms with the recession. In addition, we note the interesting pattern that Miller Lite prices dip just after the merger. We have confirmed that this is not driven by any obvious store-level outliers. It is possible that the lower Miller Lite prices were used to clear inventory so that Coors products could be incorporated into the Miller distribution channels. Figure B.1 in Appendix B does not reveal any comparable reduction in drugstore prices, however.

Table II: Changes in Retail Prices by Firm

	(i)	(ii)	(iii)	(iv)
$\mathbb{1}\{\text{MillerCoors}\} \times \mathbb{1}\{\text{Post-Merger}\}$	0.098	0.050	0.047	0.069
	(0.007)	(0.004)	(0.005)	(0.007)
$\mathbb{1}\{\text{ABI}\} \times \mathbb{1}\{\text{Post-Merger}\}$	0.087	0.040	0.038	0.062
	(0.007)	(0.005)	(0.005)	(0.007)
$\mathbb{1}\{\text{Post-Merger}\}$	-0.031	-0.007	-0.005	0.010
	(0.005)	(0.004)	(0.004)	(0.009)
log(Employment)	-	-	-0.051	0.131
	-	-	(0.080)	(0.081)
log(Earnings)	-	-	0.156	0.152
	-	-	(0.029)	(0.035)
Pre-Merger Average Price	11.75	11.14	11.14	11.14
Product Trends	No	No	Yes	Yes
Covariates	No	No	Yes	Yes
# Observations	25,740	167,695	167,695	151,525

Notes: Estimation is with OLS. The dependent variable is log real retail price. Observations are at the brand–size–region–month-year level. Column (i) includes 12 packs of Bud Light, Coors Light, Miller Lite, Corona Extra, and Heineken. Columns (ii) and (iii) includes the six-, 12, and 24 packs of these brands plus Budweiser, Michelob Light, Michelob Ultra, Coors, Miller Genuine Draft, Miller High Life, Corona Light, and Heineken Premium Light. The estimation sample spans 39 regions from 2001 to 2011, except in column (iv), which excludes June 2008 through May 2009. All regressions include a linear time trend and product (brand×size) fixed effects interacted with region fixed effects. Standard errors are clustered at the region level and shown in parentheses.

(ii) shows the results when the sample incorporates the other brands and package sizes in the sample. Absolute and relative price increases for MillerCoors and ABI are then about four and five percent, respectively. Column (iii) allows the time trend to vary freely by brand and package size and includes regional quarterly employment rates and average weekly earnings. The results are essentially unchanged. The final column excludes one year of data immediately following the merger. The estimates again reflect the pattern shown in Figure 1 and increase in magnitude. Ultimately, average prices for MillerCoors and ABI brands increased by between six and seven percent.

3.2 Cross-sectional variation in price increases

Empirical analyses of mergers sometimes assume that price effects are proportional to the predicted change in HHI (or ΔHHI) induced by the merger (e.g., Dafny, Duggan and Ramanarayanan (2012); Ashenfelter, Hosken and Weinberg (2015)). For differentiated-product markets, there is a theoretical justification for this assumption if consumers substitute to other products in proportion to their market shares, because the relevant diversion ratios can then be approximated as a function of ΔHHI.[6] In this section, we exploit cross-sectional variation in the data to evaluate whether the price patterns developed above are well explained by the region-specific ΔHHI caused by the merger.

The following regression equation specifies the log retail price of product j in region r in period t:

$$\begin{aligned} \log p^R_{jrt} &= \alpha_1 \Delta \text{HHI}_r \times \mathbb{1}\{\text{Post-Merger}\}_t \\ &+ \alpha_2 \Delta \text{MILES}_r \times \mathbb{1}\{\text{Post-Merger}\}_t \\ &+ \alpha_3 \mathbb{1}\{\text{Post-Merger}\}_t + \phi_{jr} + \tau_t + \epsilon_{jrt} \end{aligned} \tag{2}$$

where ΔHHI_r is calculated based on data from the 18 months preceding the merger (scaled to be between zero and one) and ΔMILES_r is the reduction in (thousands of) miles from the brewery to the region experienced by the Coors brands. The error structure incorporates time effects and product $\times$ region fixed effects. We estimate equation (2) separately for MillerCoors, ABI, and Modelo–Heineken.[7]

The results are shown in Table III. The price increases of MillerCoors and ABI are higher in regions with a greater ΔHHI and lower in regions that experience a greater reduction in Coors' shipping distances. This is consistent with unilateral effects theory under proportional substitution. The net effect of greater concentration and lower shipping distances is close to zero, on average.[8] Thus, the post-merger indicator accounts for most of the overall price increases shown previously; these are estimated to be 4.9 percent for MillerCoors and 4.0 percent for ABI. The magnitude and statistical significance of the post-merger indicator variable in these regressions suggest that unilateral effects may not fully account for the estimated price increases. Only weak inferences can be drawn, however, because the extent to which ΔHHI captures unilateral price effects depends on the extent to which

[6]Consider a merger that involves two products with pre-merger market shares s_j and s_k, respectively. The predicted HHI change is $\Delta\text{HHI}= 2s_js_k$. If consumer substitution is proportional to market shares, then diversion from product j to product k equals $s_k/(1-s_j)$ and can be approximated by $s_k(1+s_j)$ for small s_j. Diversion from k to j is analogous, meaning that the sum of the approximate diversion ratios is $s_j + s_k + 2s_js_k$ or, equivalently, $s_j + s_k + \Delta\text{HHI}$. We first encountered these mathematics in Shapiro (2010). Miller, Remer, Ryan and Sheu (2017) provide Monte Carlo evidence that ΔHHI is highly correlated with unilateral price effects in the specific setting of proportional substitution.

[7]This replicates the analysis of Ashenfelter, Hosken and Weinberg (2015), who estimate equation (2) with proprietary IRI data spanning 2007–2011 and 47 geographic markets. Similar results are obtained.

[8]The average increase in concentration across markets is 0.02, which is associated with a price increase of 2.0 percent ($0.997 \times 0.02 = 0.020$). The average reduction in miles is 360, which implies a price reduction of 1.5 percent ($0.36 \times -0.042 = 0.015$).

Table III: Cross-Sectional Variation in Price Increases

	Pooled	MillerCoors	ABI	Imports
ΔHHI $\times$ $\mathbb{1}$\{Post-Merger\}	0.997	1.172	1.503	-0.005
	(0.454)	(0.542)	(0.531)	(0.534)
ΔMILES $\times$ $\mathbb{1}$\{Post-Merger\}	-0.042	-0.040	-0.053	-0.028
	(0.013)	(0.016)	(0.013)	(0.014)
$\mathbb{1}$\{Post-Merger\}	0.037	0.049	0.040	0.019
	(0.012)	(0.014)	(0.013)	(0.014)
# Observations	167,695	75,315	50,810	41,570

Notes: Estimation is with OLS. The dependent variable is log real retail price. Observations are at the brand–size–region–month–year level. The estimation sample spans 39 regions from 2001 to 2011. All regressions include a linear trend and product (brand $\times$size) fixed effects interacted with region fixed effects. Standard errors are clustered at the region level and shown in parentheses.

consumer substitution is proportional to market share. This helps motivate the additional structure that we place on the model, which allows us to estimate demand elasticities from the data and account for unilateral effects directly.

3.3 Documentary record

There is documentary evidence in the public domain that supports coordinated pricing by ABI and MillerCoors. The DOJ Complaint filed to enjoin the acquisition of Grupo Modelo by ABI alleges that ABI and MillerCoors announce (nominal) price increases each year in late summer to take effect in early fall. In most geographic areas, ABI is the market share leader and announces its price increase first; in other areas, MillerCoors announces first. The price increases are usually matched by the follower and if not they are rescinded. The Complaint quotes from the normal course documents of ABI as follows:

> The specifics of ABI's pricing strategy are governed by its "Conduct Plan," a strategic plan for pricing in the United States that reads like a how-to manual for successful price coordination. The goals of the Conduct Plan include "yielding the highest level of followership in the short-term" and "improving competitor conduct over the long-term."
>
> ABI's Conduct Plan emphasizes the importance of being "Transparent – so competitors can clearly see the plan;" "Simple – so competitors can understand the 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."

The Complaint does not identify the date at which ABI adopted its Conduct Plan, but some inferences can be made from the annual reports of the companies. The 2005 SABMiller annual report describes "intensified competition" and an "extremely competitive environment." The 2005 Anheuser-Busch report states that the company was "collapsing the price umbrella by reducing our price premium relative to major domestic competitors." SABMiller characterizes price competition as "intense" in its 2006 and 2007 reports. The tenor of the annual reports changes around the time of the merger. In its 2009 report, SABMiller attributes increasing earnings before interest, taxes, and amortization expenses to "robust pricing" and "reduced promotions and discounts." In its 2010 and 2011 reports, it references "sustained price increases" and "disciplined revenue management with selected price increases."[9]

The record supports that any coordination is limited to ABI and MillerCoors. The DOJ Complaint alleges that Modelo did not join the price increases and instead adopted a "Momentum Plan" that was designed to "grow Modelo's market share by shrinking the price gaps between brands owned by Modelo and domestic premium brands." The practical consequence of the Momentum Plan is that the nominal prices of Modelo remain flat even as ABI's and MillerCoors' prices increase. This limited the ability of ABI and MillerCoors to raise prices due to the greater substitution of consumers to Modelo. The Complaint does not address the pricing practices of Heineken, though in the retail sales data we examine, the prices of Heineken's beers are similar to those of Corona.

4 Consumer demand

4.1 Model

We use the random coefficient nested logit (RCNL) model to estimate consumer demand. The RCNL model has been applied in a number of recent empirical articles (e.g., Grennan (2013); Ciliberto and Williams (2014); Conlon and Rao (2016)) and similar discrete choice random utility models have been applied to the beer industry (e.g., Hellerstein (2008); Goldberg and Hellerstein (2013); Romeo (2016); Asker (2016)).

Suppose we observe $r = 1, \ldots, R$ regions over $t = 1, \ldots, T$ time periods. There are $i = 1, \ldots, N_{rt}$ consumers in each region–period combination. Each consumer purchases one

[9] See SABMiller's Annual Report of 2005 (p. 13), 2006 (p. 5), 2007 (pp. 4 and 8), 2009 (pp. 9 and 24), 2010 (pp. 29), and 2011 (p. 28) and Anheuser-Busch's Annual Report in 2005 (p. 5). ABI's annual reports in the post-merger years are more opaque.

of the observed products ($j = 1, \ldots, J_{rt}$) or selects the outside option ($j = 0$). We refer to observed products as inside goods. The conditional indirect utility that consumer i receives from inside good j in region r and period t is

$$u_{ijrt} = x_j \beta_i^* + \alpha_i^* p_{jrt} + \sigma_j^D + \tau_t^D + \xi_{jrt} + \bar{\epsilon}_{ijrt} \tag{3}$$

where x_j is a vector of observable product characteristics, p_{jrt} is the retail price, σ_j^D allows the mean valuation of unobserved product characteristics to vary freely by product, τ_t^D allows the mean valuation of the indirect utility from consuming the inside goods to vary freely over time, ξ_{jrt} is an unobserved quality valuation specific to the region–period, and $\bar{\epsilon}_{ijrt}$ is a stochastic term.

The observable product characteristics include a constant (i.e., an indicator that equals one for an inside good), calories, package size, and an indicator for whether the product is imported. Calories is highly correlated with alcohol content and serves to distinguish the "light" beers. We control for σ_j^D and τ_t^D using product and time dummy variables, respectively. The term ξ_{jrt} is left as a structural error term. We specify the consumer-specific coefficients as $[\alpha_i^*\ , \beta_i^*]' = [\alpha, \beta]' + \Pi D_i$, where D_i is (demeaned) consumer income. The α and β parameters are the average effect of observables on indirect utility. Because the observable product characteristics are invariant over time, the mean consumer valuations for observables are absorbed by the product fixed effects in estimation.

We decompose the stochastic term using the distributional assumptions of the nested logit model, following Berry (1994), and Cardell (1997). Define two groups, $g = 0, 1$, such that group 1 includes the inside goods and group 0 the outside good. Then

$$\bar{\epsilon}_{ijrt} = \zeta_{igrt} + (1 - \rho)\epsilon_{ijrt} \tag{4}$$

where ϵ_{ijrt} is the independent and identically distributed 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 greater correlation in preferences for products of the same group and thus less consumer substitution between the inside and outside goods. To close the model, we normalize the indirect utility of the outside good such that $u_{i0rt} = \epsilon_{i0rt}$, and assume that the market sizes are 50% greater than the maximum observed unit sales within each region. The outside good includes brands outside the sample (e.g., craft beers), beer sold outside supermarkets, and non-beer beverages such as wine. Placing these products in the outside good group prompts their prices to become non-strategic in the model. Time fixed effects help control for the trend toward craft beer during the sample period.

In estimation, it is useful to decompose indirect utility such that

$$\begin{aligned} u_{ijrt} &= \delta_{jrt}(x_j, p_{jrt}, \sigma_j^D, \tau_t^D, \xi_{jrt}; \alpha, \beta) + \mu_{ijrt}(x_j, p_{jrt}, D_i; \Pi) + \zeta_{igrt} + (1-\rho)\epsilon_{ijrt} \\ \delta_{jrt} &= x_j\beta + \alpha p_{jrt} + \sigma_j^D + \tau_t^D + \xi_{jrt} \\ \mu_{ijrt} &= [p_{jrt}, x_j]' * \Pi D_i \end{aligned} \tag{5}$$

where $\delta_{jrt}(x_j, p_{jrt}, \sigma_j^D, \tau_t^D, \xi_{jrt}; \alpha, \beta)$ is the mean consumer valuation of product j in region r and period t and consumer-specific deviations are contained in $\mu_{ijrt}(x_j, D_i; \Pi) + \zeta_{igrt} + (1 - \rho)\epsilon_{ijrt}$. Suppressing function arguments, we express the market share of good j in region r and period t as

$$s_{jrt} = \frac{1}{N_{rt}} \sum_{i=1}^{N_{rt}} \frac{\exp((\delta_{jrt} + \mu_{ijrt})/(1-\rho))}{\exp(I_{igrt}/(1-\rho))} \frac{\exp I_{igrt}}{\exp I_{irt}} \tag{6}$$

where I_{igrt} and I_{irt} are the McFadden (1978) inclusive values. The ormalization on the mean indirect utility of the outside good yields $I_{i0rt} = 0$, while the inclusive value of the inside goods is $I_{i1rt} = (1-\rho)\log\sum_{j=1}^{J_{rt}} \exp((\delta_{jrt} + \mu_{ijrt})/(1-\rho))$ and the inclusive value of all goods is $I_{irt} = \log(1 + \exp I_{i1rt})$.

This RCNL model reduces to the nested logit model if $\Pi = 0$. This yields an equation that is linear in its parameters:

$$\log(s_{jrt}) - \log(s_{0rt}) = x_j\beta + \alpha p_{jrt} + \sigma_j^D + \tau_t^D + \rho\log(\overline{s}_{jrt|g}) + \xi_{jrt} \tag{7}$$

where $\overline{s}_{jrt|g} = s_{jrt}/\sum_{j=1}^{J_{rt}} s_{jrt}$ is the conditional share of product j among the inside goods. In this formulation, the nesting parameter ensures that the estimated elasticities are not overly sensitive to the market size assumption. The model nonetheless retains the property that substitution patterns among inside goods are a function only of market shares. The full RCNL model relaxes this restriction by allowing consumer income to affect relative choice probabilities. It also allows the recession to affect demand in a natural way.

4.2 Estimation and instruments

We estimate the demand model using the nested fixed point procedure of Berry, Levinsohn and Pakes (1995). This approach derives a generalized method of moments (GMM) estimator from the population moment condition $E[Z' \cdot \omega(\theta_0^D)] = 0$, where $\omega(\cdot)$ is a vector defined below, $\theta_0^D = (\alpha, \Pi, \rho)$ is the vector of population parameters, and Z is a conformable matrix of instruments. The GMM estimate is

$$\widehat{\theta}^D = \arg\min_\theta \omega(\theta)' Z A^{-1} Z' \omega(\theta) \tag{8}$$

for some positive definite weighting matrix A. For any set of candidate parameters $(\tilde{\Pi}, \tilde{\rho})$, a contraction mapping identifies the mean utility levels that equate the observed and predicted market shares. Formally, we obtain a vector $\delta^*(x, p, S; \tilde{\Pi}, \tilde{\rho})$ as the solution to the implicit system of equations $s(x, p, \delta^*; \tilde{\Pi}, \tilde{\rho}) = S$, where $s(\cdot)$ is a vector of market shares defined by equation (6), p is a vector of prices, and S is a vector of observed market shares. The vector $\omega(\tilde{\Pi}, \tilde{\rho})$ is the residual from a two-stage least squares (2SLS) regression of $\delta^*(x, p, S; \tilde{\Pi}, \tilde{\rho})$ on price and the fixed effects. This recovers the structural error term if evaluated at the population parameters (i.e., $\xi_{jrt} = \omega_{jrt}(\Pi, \rho)$). The price coefficient in the 2SLS regression is an estimate of α, which allows us to restrict the nonlinear search to Π and ρ.

We employ the standard two-step procedure for GMM estimation (Hansen (1982)). In the first step, we set $A = Z'Z$. In the second step, we reestimate the model using an optimal weighting matrix that employs an Eicker–White–Huber cluster correction to correct for heteroskedasticity, autocorrelation, and within-region correlations. Asymptotic consistency is obtained as the number of regions increases.

Identification requires at least one instrument for price and each nonlinear parameter. Prices are likely to be correlated with the structural error term because firms set prices with knowledge of product- and market-specific consumer valuations. This creates a standard endogeneity problem. Further, as highlighted by Berry and Haile (2014), the presence of heterogeneity in consumer preferences for product characteristics introduces a simultaneity problem that arises from the interaction of unknown demand parameters with market shares. This is because the mean utilities that equate observed shares to predicted shares depend on the parameters that govern how consumer heterogeneity determines choices: $\delta^*(x, p, S; \tilde{\Pi}, \tilde{\rho}) - \alpha p_{jtr} - \sigma_j - \sigma_t = \xi_{jrt}$. In our specification of the RCNL, this heterogeneity is due to the income terms (Π) and the nested logit term (ρ).[10]

The first set of instruments that we use address the endogeneity of prices. It includes the distance between the brewery and the region (miles $\times$ diesel index) and an indicator equal to one for ABI and MillerCoors products after the merger. Both instruments arise from the supply side of the model; distance shifts marginal costs and the indicator captures a change in the competitive structure of the industry. The relevance of the indicator is

[10]Appendix D.1 provides more detail on how consumer heterogeneity in preferences for product characteristics creates endogeneity problems. See also Berry and Haile (2014).

suggested by the observed price increases after the Miller/Coors merger. Given the time and product fixed effects, the indicator is valid if the changes in the structural error terms of ABI and MillerCoors, before versus after the merger, are not systematically different from the changes in the structural error terms of Modelo and Heineken.

The second set of instruments helps identify the nested logit parameter, which governs the degree of correlation in unobserved preferences for the inside goods. What is required is exogenous variation in the conditional shares of the inside goods (i.e., variation in $\overline{s}_{jrt|g} = s_{jrt} / \sum_{j=1}^{J_{rt}} s_{jrt}$). This is made clear in the linear formulation of the nested logit model in equation (7). We use as instruments the number of products in the market and the distance summed across all products in the market. The effect of these variables on choice probabilities need not be uniform in the sample and, to add flexibility, we incorporate interactions with indicators for ABI and Miller/Coors products. The number of products is a standard instrument and should be negatively correlated with the conditional share. Total distance captures variation in the marginal costs of competing products and should be positively correlated with conditional share. Validity of this instrument requires the structural error term to be uncorrelated with the number of products.

Finally, the parameters governing consumer heterogeneity in preferences for characteristics are identified by the correlation between local demographics and product market shares. To identify these parameters, we use mean income interacted with the observed product characteristics (a constant, calories, package size, and an import dummy), which provide the requisite variation. Under the assumption that the structural error term is mean independent of income and product characteristics, these instruments are valid. Romeo (2014) provides evidence that similar instruments improve the numerical performance of random coefficient logit estimates. There are 12 instruments in total. We evaluate the relevance of these instruments, along with related considerations, in Appendix D.1.

4.3 Results of demand estimation

Table IV presents the results of demand estimation. Column (i) corresponds to the nested logit demand model, which we estimate with 2SLS to provide a simple benchmark. The remaining columns correspond to the full RCNL model. We use two main specifications. In columns (ii) and (iii), consumer income affects preferences for price, the inside good constant, and calories. In columns (iv) and (v), income affects preferences for the constant, calories, imports, and package size. Both specifications break the logit substitution patterns between domestic and imported beers, albeit with different mechanisms. The units of observation are

Table IV: Baseline Demand Estimates

Demand Model:		NL-1	RCNL-1	RCNL-2	RCNL-3	RCNL-4
Data Frequency:		Monthly	Monthly	Quarterly	Monthly	Quarterly
Variable	Parameter	(i)	(ii)	(iii)	(iv)	(v)
Price	α	-0.1312	-0.0887	-0.1087	-0.0798	-0.0944
		(0.0884)	(0.0141)	(0.0163)	(0.0147)	(0.0146)
Nesting Parameter	ρ	0.6299	0.8299	0.7779	0.8079	0.8344
		(0.0941)	(0.0402)	(0.0479)	(0.0602)	(0.0519)
Demographic Interactions						
Income×Price	Π_1		0.0007	0.0009		
			(0.0002)	(0.0003)		
Income×Constant	Π_2		0.0143	0.0125	0.0228	0.0241
			(0.0051)	(0.0055)	(0.0042)	(0.0042)
Income×Calories	Π_3		0.0043	0.0045	0.0038	0.0031
			(0.0016)	(0.0017)	(0.0018)	(0.0015)
Income×Import	Π_4				0.0039	0.0031
					(0.0019)	(0.0016)
Income×Package Size	Π_5				-0.0013	-0.0017
					(0.0007)	(0.006)
Other Statistics						
Median Own Price Elasticity		-3.81	-4.74	-4.33	-4.45	-6.10
Median Market Price Elasticity		-1.10	-0.60	-0.72	-0.60	-0.69
Median Outside Diversion		29.80%	12.96%	16.98%	13.91%	11.82%
J-Statistic			13.94	13.75	13.91	14.15

Notes: This table shows the baseline demand results. We use 2SLS for estimation in column (i) and GMM in columns (ii) to (v). There are 94,656 observations at the brand–size–region–month–year level in columns (i), (ii), and (iv) and 31,784 observations at the brand–size–region–year–quarter level in columns (iii) and (v). The samples exclude the months/quarters between June 2008 and May 2009. All regressions include product (brand×size) and period (month or quarter) fixed effects. The elasticity and diversion numbers represent medians among all the brand–size–region–month/quarter–year observations. Standard errors are clustered by region and shown in parentheses.

brand–size–region–year–month combinations in columns (i), (ii), and (iv) and brand–size–region–year-quarter combinations in columns (iii) and (v). All regressions include product (i.e., brand×size) and time fixed effects.

The coefficients are precisely estimated and have the expected signs. The median own-price elasticities range from -4.33 to -6.10 for the RCNL models.[11] The market price elasticities are much lower, indicating that most substitution occurs within inside goods, rather than between the inside goods and the outside good. This can be recast in terms

[11]The own-price elasticities are somewhat greater than the own-price elasticities reported by Romeo (2016) and somewhat smaller than those reported by Hellerstein (2008). Most similar are the elasticities of Slade (2004) and Pinske and Slade (2004), obtained for the U.K. beer industry.

of diversion: the outside good is the second-best choice for 12%-17% of consumers. The interactions reveal that consumers with higher incomes are less sensitive to price and tend to prefer the inside goods, imported brands, more calories, and smaller package sizes. Whether the monthly or quarterly data are used in estimation matters little. Finally, the Sargan–Hansen J-statistics are asymptotically χ^2 distributed with either eight (columns (ii) and (iii)) or seven (columns (iv) and (v)) degrees of freedom under the null hypothesis that the models are valid. The models cannot be rejected at the 95% confidence level.

Table V presents more detail on the elasticities that arise in the RCNL-1 specification. We provide the full elasticity matrix for 12 packs, along with aggregated cross-elasticities that summarize substitution from the 12 packs to selected categories of beer. One noticeable pattern is that own price elasticities tend to be somewhat higher for more expensive products. (This is not imposed in RCNL-1 due to the income×price interaction.) The logit restriction that consumers substitute to other products in proportion to their market shares is substantially relaxed. This can be seen by observing the heterogeneity that exists within a single column. For instance, column 1 shows that consumers of Bud Light 12 packs substitute disproportionately toward similar beers such as Budweiser, Coors Light, and Miller Lite and also toward the larger package sizes. Column (5) shows that, by contrast, consumers of Corona Extra substitute disproportionately toward Heineken and smaller package sizes.

Table V: Mean Elasticities for 12 Pack Products from Specification RCNL-1

	Brand/Category	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
							Product-Specific Own and Cross-Elasticities							
(1)	Bud Light	-4.389	0.160	0.019	0.182	0.235	0.101	0.146	0.047	0.040	0.130	0.046	0.072	0.196
(2)	Budweiser	0.323	-4.272	0.019	0.166	0.258	0.103	0.166	0.047	0.039	0.121	0.043	0.068	0.183
(3)	Coors	0.316	0.154	-4.371	0.163	0.259	0.102	0.167	0.046	0.038	0.119	0.042	0.066	0.180
(4)	Coors Light	0.351	0.160	0.019	-4.628	0.230	0.100	0.142	0.047	0.041	0.132	0.047	0.073	0.199
(5)	Corona Extra	0.279	0.147	0.018	0.137	-5.178	0.108	0.203	0.047	0.035	0.104	0.035	0.061	0.158
(6)	Corona Light	0.302	0.151	0.018	0.153	0.279	-5.795	0.183	0.048	0.037	0.113	0.039	0.065	0.171
(7)	Heineken	0.269	0.145	0.018	0.131	0.311	0.108	-5.147	0.047	0.035	0.101	0.034	0.059	0.153
(8)	Heineken Light	0.240	0.112	0.014	0.124	0.210	0.086	0.138	-5.900	0.026	0.089	0.028	0.051	0.135
(9)	Michelob	0.301	0.140	0.015	0.146	0.208	0.089	0.135	0.042	-4.970	0.116	0.036	0.061	0.175
(10)	Michelob Light	0.345	0.159	0.019	0.181	0.235	0.101	0.146	0.047	0.041	-5.071	0.046	0.072	0.196
(11)	Miller Gen. Draft	0.346	0.159	0.019	0.182	0.235	0.101	0.146	0.047	0.040	0.130	-4.696	0.072	0.196
(12)	Miller High Life	0.338	0.159	0.019	0.177	0.242	0.102	0.153	0.047	0.040	0.127	0.045	-3.495	0.191
(13)	Miller Lite	0.344	0.159	0.019	0.180	0.237	0.101	0.148	0.047	0.040	0.129	0.046	0.071	-4.517
(14)	Outside Good	0.016	0.007	0.001	0.009	0.011	0.005	0.006	0.002	0.002	0.006	0.002	0.003	0.009
							Cross Elasticities by Category							
	6 Packs	0.307	0.152	0.018	0.155	0.275	0.104	0.180	0.047	0.038	0.115	0.039	0.065	0.174
	12 Packs	0.320	0.154	0.019	0.163	0.250	0.102	0.161	0.047	0.039	0.121	0.042	0.068	0.183
	24 Packs	0.356	0.160	0.019	0.189	0.222	0.099	0.136	0.047	0.041	0.134	0.048	0.073	0.201
	Domestic	0.349	0.160	0.019	0.184	0.229	0.100	0.142	0.047	0.040	0.131	0.047	0.072	0.197
	Imported	0.279	0.147	0.018	0.138	0.301	0.108	0.200	0.047	0.035	0.104	0.035	0.061	0.158

Notes: This table provides the mean elasticities of demand for 12 packs based on the RCNL-1 specification (column (ii) of Table IV). The cell in row i and column j is the percentage change in the quantity of product i with respect to the price of product j. Means are calculated across year–month–region combinations. The category cross-elasticities are the percentage change in the combined shares of products in the category due to a 1 percent change in the price of the product in question. Letting the category be defined by the set B, we calculate $\left(\sum_{j\in B, j\neq k} \frac{\partial s_j(p)}{\partial p_k}\right) \frac{p_k}{\sum_{j\in B, j\neq k} s_j(p)}$. The categories exclude the product in question. Thus, for example, the table shows that a 1 percent change in the price of a Bud Light 12 pack increases the sales of other 12 packs by 0.320 percent.

5 Supply

5.1 Model

We estimate a model of differentiated products price competition in which ABI and MillerCoors to partially or fully internalize their pricing externalities in the post-merger periods. Brewers in the model sell directly to consumers; a more sophisticated treatment of the retail sector is provided in Appendix E. The vector of equilibrium prices in each region–period satisfies the first-order condition

$$p_t = mc_t - \left[\Omega_t(\kappa) \circ \left(\frac{\partial s_t(p_t;\theta^D)}{\partial p_t}\right)^T\right]^{-1} s_t(p_t;\theta^D) \tag{9}$$

where Ω_t is the ownership matrix, s_t is a vector of market shares, and the operation $\circ$ is element-by-element matrix multiplication. We suppress region subscripts for brevity. The (j, k) element of the ownership matrix equals one if products j and k are produced by the same firm. The (j, k) element equals κ if products j and k are sold by ABI and MillerCoors and the period postdates the merger. Otherwise the (j, k) element equals zero. This generates Nash–Bertrand competition in the post-merger periods if $\kappa = 0$ and joint profit maximization for ABI and MillerCoors if $\kappa = 1$.[12]

To illustrate, consider a hypothetical region in a pre-merger period t_1^* and a post-merger period t_2^*, and suppose that there are $j = 1, ...4$ products sold by ABI, Miller, Coors, and Modelo, respectively. The ownership matrices are given by

$$\Omega_{t_1^*} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad \Omega_{t_2^*} = \begin{bmatrix} 1 & \kappa & \kappa & 0 \\ \kappa & 1 & 1 & 0 \\ \kappa & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{10}$$

The pre-merger ownership matrix at t_1^* is diagonal because competition is Nash–Bertrand and each firm sells a single product in the hypothetical region. The post-merger matrix reflects that Miller and Coors fully internalize how their prices affect each other and the κ parameter dictates the extent to which ABI and MillerCoors internalize the effect of their

[12] Under certain assumptions the parameter κ can be interpreted as a conduct parameter (Black, Crawford, Lu and White (2004) and Sullivan (2016)). As Corts (1999) notes, if the true model implies variation in κ over time within each regime this interpretation is problematic. Nevertheless, a finding that κ is statistically different than zero provides a test for post-merger Nash–Bertrand competition. See Bresnahan (1989), particularly Section 3.4, and Porter (1983) for a discussion of identification with regime shifts.

prices. An important restriction is that Modelo and Heineken price *à la* Nash, which we motivate on the basis of qualitative evidence discussed earlier.

To complete the supply-side model, we parameterize the marginal cost of product j in region r and period t as follows:

$$mc_{jrt} = w_{jrt}\gamma + \sigma_j^S + \tau_t^S + \mu_r^S + \eta_{jrt}, \tag{11}$$

where w_{jrt} is a vector that includes the distance (miles $\times$ diesel index) between the region and brewery and an indicator for MillerCoors products in post-merger periods. This allows the Miller/Coors merger to affect marginal costs through the rationalization of distribution and through residual cost synergies unrelated to distance. Unobserved costs depend on the product, region, and period-specific effects, σ_j^S, μ_c^S, and τ_t^S, which we control for with fixed effects, as well as on η_{jrt}, which we leave as a structural error term.[13]

5.2 Estimation and instruments

We estimate the supplyside of the model taking as given the demand results. We include the supply-side parameters to be estimated in the vector $\theta_0^S = (\kappa, \gamma)$. For each candidate parameter vector $\widetilde{\theta}^S$, we calculate the markups and observed marginal costs and obtain the structural error as a function of the parameters:

$$\eta_{rt}^*(\widetilde{\theta}^S; \widehat{\theta}^D) = p_{rt} - w_{jrt}\widetilde{\gamma} - \sigma_j^S - \tau_t^S - \mu_r^S - \left[\Omega_t(\widetilde{\kappa}) \circ \left(\frac{\partial s_t(p_{rt}; \widehat{\theta}^D)}{\partial p_{rt}}\right)^T\right]^{-1} s_t(p_{rt}; \widehat{\theta}^D) \tag{12}$$

Identification rests on the population moment condition $E[z' \cdot \eta^*(\theta_0^S)] = 0$, where $\eta^*(\theta_0^S)$ is a stacked vector of structural errors and z is a conformable vector that contains an excluded instrument. The method-of-moments estimate is

$$\widehat{\theta}^S = \arg\min_\theta \eta^*(\theta; \widehat{\theta}^D)' z z' \eta^*(\theta; \widehat{\theta}^D) \tag{13}$$

[13] The slope of the marginal cost function influences the magnitude of price changes that arise from unilateral effects. Suppose that ABI has an upward-sloping marginal cost function. Then, as consumers shift to ABI in response higher prices from MillerCoors, ABI has both demand- and cost-side incentives to raise prices. Because the identification strategy is based on whether prices differ from what would be predicted on the basis of unilateral effects (accounting for changes in demand/costs), the estimated κ parameter would be biased upward unless the specification accounts for increasing costs. We assume a constant marginal cost function. There is little evidence that ABI experienced capacity constraints over the sample period.

We concentrate the fixed effects and the marginal cost parameters out of the optimization problem using OLS to reduce the dimensionality of the nonlinear search. We cluster the standard errors at the region level and make an adjustment to account for the incorporation of demand-side estimates (Wooldridge (2010)). Details are provided in Appendix D.

The markup term in equation (12) is endogenous because unobserved costs enter implicitly through price. We instrument with an indicator that equals one for ABI and MillerCoors in the post-merger periods. The power of the instrument is supported by the descriptive regression results. Validity holds if the unobserved costs of ABI are orthogonal to the instrument. Given the specification of the marginal cost function, this will be the case if changes in the unobserved costs of ABI, before versus after the merger are not systematically different from changes in the unobserved costs of Modelo and Heineken. This is because the product and time fixed effects absorb level effects in the marginal cost function. Further, the MillerCoors post-merger indicator allows the merger to shift the marginal costs of MillerCoors and thus isolates the comparison between ABI and Modelo–Heineken.

We comment briefly on the empirical variation that identifies the coefficients. The κ estimate is positive if the post-merger prices of ABI exceed what can be explained by the unilateral effects of the Miller/Coors merger. However, a positive κ affects the prices of both ABI and MillerCoors. Thus, the post-merger MillerCoors parameter in the marginal cost function is negative if post-merger MillerCoors prices are lower than what can be explained by unilateral effects and the κ estimate. Consider a simple numerical example. Demand is logit and competition is initially Nash–Bertrand among three symmetric firms. The prices are 1.00, marginal costs are 0.70, and market shares are 0.20. The share of the outside good is 0.40. This is sufficient to calibrate the demand system. Allow the first two firms to merge. Prices under post-merger Nash–Bertrand competition are (1.06,1.06,1.01). If instead $\kappa = 0.50$ then the post-merger prices are (1.10, 1.10, 1.07). If $\kappa = 0.50$ and the merging firms reduce their marginal costs to 0.50 then post-merger prices are (1.09, 1.09, 1.08). The way the prices of ABI and MillerCoors change with the merger drives the estimates of the supply-side parameters.

5.3 Supply-side estimation results

Table VI presents the supply-side results. As described, each column corresponds to one of the baseline demand specifications. The marginal cost functions incorporate product (i.e., brand$\times$size), period (month or quarter), and region fixed effects in all cases. As shown, the estimates of κ are positive and statically significant. The null of Nash–Bertrand competition

Table VI: Baseline Supply Estimates

Demand Model:		NL-1	RCNL-1	RCNL-2	RCNL-3	RCNL-4
Data Frequency:		Monthly	Monthly	Quarterly	Monthly	Quarterly
Variable	Parameter	(i)	(ii)	(iii)	(iv)	(v)
Post-Merger Internalization	κ	0.374	0.264	0.249	0.286	0.342
of Coalition Pricing Externalities		(0.034)	(0.073)	(0.087)	(0.042)	(0.054)
Marginal Cost Parameters						
MillerCoors×PostMerger	γ_1	-0.608	-0.654	-0.649	-0.722	-0.526
		(0.039)	(0.050)	(0.060)	(0.042)	(0.040)
Distance	γ_2	0.142	0.168	0.163	0.169	0.148
		(0.046)	(0.059)	(0.059)	(0.060)	(0.049)

Notes: This table shows the baseline supply results. We use the method of moments for estimation. There are 94,656 observations at the brand–size–region–month–year level in columns (i), (ii), and (iv) and 31,784 observations at the brand–size–region–year–quarter level in columns (iii) and (v). The samples exclude the months/quarters between June 2008 and May 2009. All regressions include product (brand×size), period (month or quarter), and region fixed effects. Standard errors clustered by region and shown in parentheses.

in the post-merger periods is easily rejected. With the RCNL demand specifications, the estimates range from 0.249 to 0.342. Strictly interpreted, this corresponds to ABI and MillerCoors internalizing between roughly a quarter and a third of their price effects on the other's profits in the post-merger periods.

Brewer markups can be obtained from the κ estimates and the structure of the model.[14] Table VII provides the average markup for each product in the data both before and after the Miller/Coors merger, based on the RCNL-1 specification (column (ii) in Tables IV and VI). Across all 94,656 brand–size–month–region observations, the average markup is \$3.60 on an equivalent-unit basis, which accounts for 34% of the retail price. The average markups on ABI 12 packs tend to be about \$0.70 higher in the post-merger periods. This result reflects the higher retail prices previously shown in Figure 1. The markups on Miller 12 packs increase by \$1.40 and the markups on Coors products increase by \$1.80. Those changes can be attributed to the combined impact of higher retail prices and lower marginal costs. The markups on imported beers do not change much over the sample period.

Turning to the marginal cost shifters, we find the estimated distance parameters range from 0.148 to 0.169 with the RCNL demand models. The magnitude of the coefficient indicates that marginal costs that scale with shipping distances account for 2–3% of the retail price, on average. Total distribution costs may be partially absorbed by the fixed

[14]Equation (9) allows for the retail price to be decomposed into brewer markups and marginal costs. We show in Appendix E that the magnitude of marginal costs is sensitive to the incorporation of retail market power, which has an effect that is economically similar to a per-unit tax that must be paid by the brewers. The brewer markups are largely unaffected by the incorporation of retail market power.

Table VII: Brewer Markups from RCNL-1

	6 Packs		12 Packs		24 Packs	
Brand	Pre	Post	Pre	Post	Pre	Post
Bud Light	3.63	4.34	3.52	4.24	3.43	4.13
Budweiser	3.79	4.49	3.66	4.38	3.55	4.25
Coors	2.70	4.39	2.56	4.31	2.44	4.18
Coors Light	2.47	4.21	2.36	4.14	2.28	4.04
Corona Extra	3.30	3.18	3.04	2.91	3.04	3.03
Corona Light	3.02	2.91	2.75	2.65	2.87	2.80
Heineken	3.20	3.14	2.98	2.92	3.22	3.33
Heineken Light	2.87	2.81	2.61	2.50	2.75	2.69
Michelob	3.69	4.47	3.62	4.38	3.34	4.28
Michelob Light	3.61	4.34	3.53	4.23	3.46	4.06
Miller Gen. Draft	2.89	4.26	2.77	4.16	2.68	4.09
Miller High Life	2.91	4.28	2.80	4.20	2.74	4.13
Miller Lite	2.89	4.25	2.78	4.18	2.69	4.07

Notes: This table provides the average markups for each brand–size combination separately for the pre-merger and post-merger periods, based on the RCNL-1 specification shown in column (ii) of Tables IV and VI.

effects; Tremblay and Tremblay (2005, p. 162) peg taxes and shipping at 17% of the retail price in 1996. The marginal cost specification allows for the Miller/Coors merger to produce efficiencies through both a reduction in shipping distance and a downward shift in marginal costs common to all regions. The estimates of the latter effect range from \$0.66 to \$0.70 with the RCNL demand models. This result most likely reflects distributional savings that are not captured in the distance between the brewery and retailer location. Our estimates imply a reduction in the marginal cost of Coors Light of about 14%, which can be compared against the 11% reduction predicted in the trade press (e.g., van Brugge et al (2007)).

Figure 2 explores the cross-sectional variation in the marginal cost reductions due to the Miller/Coors merger based on the RCNL-1 specification. Scatter plots are shown for Coors Light 12 packs (left panel) and Miller Lite 12 packs (right panel). The horizontal dimension shows the change in marginal cost due to the merger. The vertical dimension is the corresponding change in the retail price, which we obtain by recomputing the equilibrium under the counterfactual that the merger does not reduce marginal costs. The cost reductions for Coors Light range from \$0.60 to \$1.30, depending on the region. The average pass-through is about a \$0.80 price reduction per \$1.00 cost reduction, and similar pass-through arises for other products. Miller Lite displays less cross-sectional variation due to the more limited distance reductions. The results are broadly consistent with those of Ashenfelter, Hosken, and Weinberg (2014b), who find that shipping efficiencies reduce prices by 1.8% in the average market. By point of comparison, a counterfactual simulation in which we allow

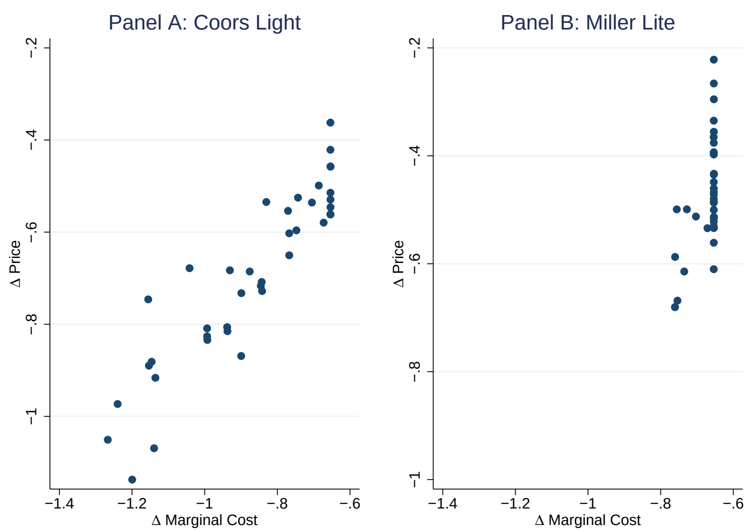

Figure 2: Change in Price against Change in Marginal Cost

Notes: This figure plots the average regional difference in counterfactual price with no efficiencies and the observed price against the average regional difference in marginal cost for 12 packs of Coors Light and Miller Lite. Each dot is a region average in 2011 and is based on the RCNL-1 specification.

the merger to change costs only through shipping distance reductions implies prices that are 2.4% lower than they would be with the initial distances.

We now discuss two subjects related to supply-side identification. First, if post-merger competition is Nash–Bertrand, then higher post-merger marginal costs of ABI would be needed to rationalize the observed prices, relative to the marginal costs implied by the baseline model. To quantify what would be required to explain the data, we obtain marginal costs under two specific supply-side assumptions: (i) competition is Nash–Bertrand in every period and (ii) pre-merger competition is Nash–Bertrand and post-merger competition features partial coordination, as implied by our baseline results. We fit the following equation to the data using OLS separately for each ABI brand:

$$\begin{aligned} \log cost_{mjrt} &= \beta_{1j}\mathbb{1}\{\text{Post-Merger}\}_t \\ &+ \beta_{2j}\mathbb{1}\{\text{Bertrand}\}_m \times \mathbb{1}\{\text{Post-Merger}\}_t \\ &+ \phi_{jr} + \tau_j \times t + \epsilon_{jrt} \end{aligned} \tag{14}$$

where $cost_{mjrt}$ is the inferred marginal cost of producing product j in region r in period t under model m. The indicator $\mathbb{1}\{\text{Post-Merger}\}_t$ equals one for the post-merger periods and $\mathbb{1}\{\text{Bertrand}\}_m \times \mathbb{1}\{\text{Post-Merger}\}_t$ is an interaction with an indicator that equals one if

Table VIII: Changes in ABI Log Costs with Bertrand and Coordination

	Budweiser	Bud Light	Michelob Light	Michelob Ultra
$\mathbb{1}\{\text{Post-Merger and Bertrand}\}$	0.122	0.120	0.089	0.102
	(0.006)	(0.006)	(0.004)	(0.007)
$\mathbb{1}\{\text{Post-Merger}\}$	0.016	-0.002	-0.044	0.050
	(0.014)	(0.011)	(0.011)	(0.013)

Notes: The dependent variable is log marginal costs from (i) the baseline model and (ii) an alternative with Nash–Bertrand pricing in all periods. The RCNL-1 specification is used to obtain the implied marginal costs. Observations are thus at the brand–size–region–month–scenario level. The estimation sample excludes observations from June 2008 through May 2009. All regressions include product (brand×size) fixed effects interacted with region fixed effects. Standard errors are clustered at the region level and shown in parentheses.

the observation is generated from the Bertrand model. The models are identical in the pre-merger periods, so there is no need to include the non-interacted $\mathbb{1}\{\text{Bertrand}\}_m$ indicator. The specification includes product×model fixed effects and product-specific trends.[15]

Table VIII presents the results. The first column indicates that it would take approximately 13.8% ($12.2 + 1.6 = 13.8$) increase in the costs of Budweiser relative to pre-merger costs to rationalize observed prices under Nash–Bertrand competition. By contrast, the baseline model implies that the marginal costs of ABI increase by 1.6% with the Miller/Coors merger. The results are similar for the other ABI brands. There is some documentary evidence in the public domain that helps assess the plausibility of the cost predictions. InBev purchased Anheuser-Busch in November 2008. It revised the pay system, ended pension contributions and life insurance for retirees, and transferred the foreign beer operations of Anheuser-Busch to Inbev (Ascher (2012)). None of these changes affect distribution costs. Thus, it seems likely that ABI's marginal costs were flat after the Miller/Coors merger. Additionally, the implied ABI cost increases under Nash–Bertrand competition are of similar magnitude to the cost reductions we estimate for MillerCoors, which makes us skeptical that they would pass without notice in the annual reports of ABI and the popular press.[16]

Second, we emphasize that the κ parameter allows us to test for a *change* in the equilibrium concept. Pre-merger competition is normalized to Nash–Bertrand in the baseline model, but other normalizations are possible. Table IX reports estimates obtained under some of these alternatives. Specifically, we impose a non-zero pre-merger κ parameter (0.10, 0.20, …, 0.50) that governs interactions between the domestic brewers and we estimate

[15] The exercise is in the spirit of Bresnahan (1987).

[16] While we cannot rule out that the new management of ABI is simply more inclined to coordinate, it is worth noting that the price increases of Figure 1 predate the close of the InBev/ABI merger.

Table IX: Supply-Side Estimates with Alternative Pre-Merger Normalizations

	(i)	(ii)	(iii)	(iv)	(v)
Post-Merger Internalization of Coalition Pricing Externalities	0.320 (0.066)	0.382 (0.058)	0.448 (0.051)	0.518 (0.043)	0.593 (0.036)
Pre-Merger Internalization	0.10	0.20	0.30	0.40	0.50

Notes: This table shows the baseline supply results based on method of moments estimation. The results are generated with the RCNL-1 demand specification. There are 94,656 observations at the brand–size–region–month–year level. The sample excludes the months between June 2008 and May 2009. All regressions include the baseline marginal cost shifters, as well as product and time fixed effects. Standard errors are clustered by region and shown in parentheses.

the corresponding post-merger parameters. As shown, the higher the pre-merger parameter, the higher the post-merger parameter. In each case, the post-merger parameter is statistically different from the pre-merger normalization, so the null hypothesis of no change in the equilibrium concept can be rejected. This reinforces that identification hinges on whether observed price changes can be explained by the supply-side model, holding fixed the equilibrium concept. The nature of pre-merger competition is not determined

5.4 Coordination and the Great Recession

One limitation of the baseline econometric results is that they are consistent with post-merger coordination but do not directly inform whether this is actually caused by the merger. This mirrors the documentary record summarized in Section 3.3, which suggests softer price competition after 2008 but does not explain why the shift occurred. It is difficult to draw strong conclusions based on the evidence about why coordination became more feasible or more effective after the Miller/Coors merger. The merger is coincident with the Great Recession, so it is worth exploring the alternative explanation that coordination is supported by weak demand (e.g., Rotemberg and Saloner (1986)). The recession appears to have had adverse effects on ABI and MillerCoors: unit sales decrease in both absolute terms and relative to Modelo–Heineken, and ABI's 2009 annual report (p. 17) refers to "an economic environment that was the most difficult our industry has seen in many years."

Some empirical evidence belies this alternative hypothesis. The prices of ABI and MillerCoors continue to increase relative to Modelo/Heineken over 2009–2011 during a period of macroeconomic recovery and are positively associated with household earnings (see Section 3). Further, we estimate a supply-side model that allows for a different κ parameter in each post-merger period. Figure 3 plots the point estimates along with a 95% confidence interval.

The estimates increase over time.[17] It is also possible to allow κ to vary with mean income, thereby exploiting the cross-sectional variation. To do so, we constrain the values to be between zero and one, as follows:

$$\kappa_{rt} = \frac{\exp(a \times \mathbb{1}\{\text{Post-Merger}\}_{rt} + b \text{ Mean Income}_{rt})}{1 + \exp(a \times \mathbb{1}\{\text{Post-Merger}\} + b \text{ Mean Income}_{rt})} \tag{15}$$

where a and b are parameters and Mean Income is demeaned so that its average is zero. With the RCNL-3 specification, we estimate a to be 0.456 (standard error of 0.088) and b to be 0.033 (standard error of 0.039), which again does not support countercyclical coordination. The average values of κ are 0.50 and 0.61 in the pre-merger and post-merger periods, respectively, and the standard deviations are 0.05 and 0.04. The results with the RCNL-1 specification are noisier: We estimate a to be 0.475 (standard error of 0.641) and b to be -0.198 (standard error of 1.024). Given the totality of the evidence, we view the merger as the most likely catalyst of softer price competition.

6 Counterfactual analysis

We use counterfactual analysis to study the mechanisms through which the Miller/Coors merger affects market outcomes. We focus on (i) unilateral effects resulting from the internalization of competition between Miller and Coors, (ii) coordinated effects that we capture with the non-zero post-merger κ parameter, and (iii) marginal cost reductions from the merger efficiencies. Our analysis suggests that the raw data reflect coordinated effects and cost reductions. Therefore, this is our interpretation of the data. We recompute equilibrium with the RCNL-1 specification under the following four counterfactual scenarios:

- The merger does not occur.
- The merger occurs with efficiencies and without coordinated effects.
- The merger occurs without efficiencies and without coordinated effects.
- The merger occurs without efficiencies and with coordinated effects.

[17] Coordinated effects could increase over time for a variety of reasons. One plausible explanation is that consumer expectations for prices are slow to adjust, which causes some price stickiness (in nominal terms). While any such friction could also apply to alternative explanations in which coordination is triggered by an event coincident with the merger, Figure 3 remains inconsistent with the model of Rotemberg and Saloner, which predicts countercyclical coordination.

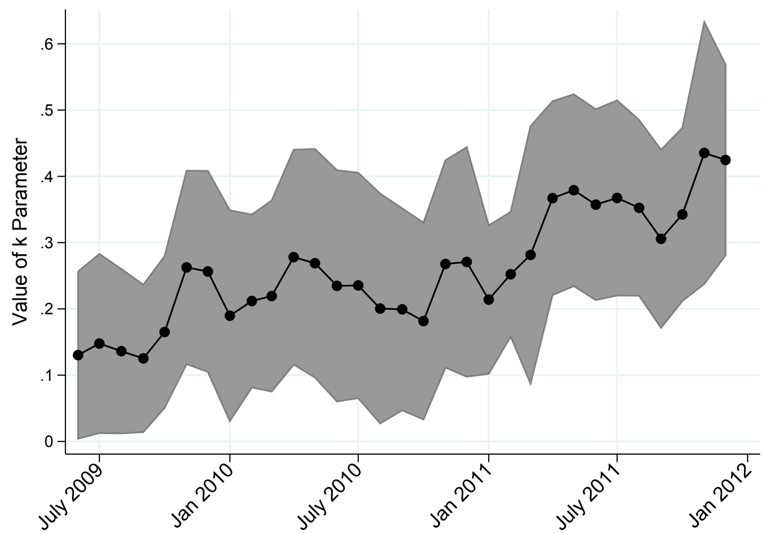

Figure 3: Time-Varying Estimates of the κ Parameter

Notes: This figure plots the point estimates of κ for each post-merger month–year combination, along with a 95% confidence interval. Estimation is carried out with the method of moments and we use the results generated with the RCNL-1 demand specification. There are 94,656 observations at the brand–size–region–month–year level. The sample excludes the months between June 2008 and May 2009. The marginal cost function includes the baseline cost shifters and product (i.e., brand$\times$size), period, and region fixed effects. The instruments include the interaction of an indicator for ABI products with indicators for each post-merger time period. The confidence interval is calculated with standard errors clustered by region.

Figure 4 shows the prices of Miller Lite 12 packs under each of the five scenarios. Prices in the "No Merger" scenario are substantially lower than in the raw data and appear to roughly track the pre-merger trend. The unilateral effects and marginal cost reductions roughly offset each other, so that prices in the "No Merger" and "Unilateral, Efficiencies" scenarios are quite similar. The magnitudes of the unilateral effects and the efficiencies are both large, as evidenced by the higher prices that arise in the "Unilateral, No Efficiencies" scenario. Finally, the coordinated effects are strong enough that prices increase even given the large marginal cost reductions: The merger increases prices by about 50 cents, on average, per 12 pack, relative to the "No Merger" scenario.

Figure 5 shows the prices of Bud Light 12 packs. We omit the "Unilateral, Efficiencies" scenario from the graph because they are nearly exactly the same as the "No Merger" scenario. (This is because unilateral effects were almost entirely offset by efficiencies for Miller and Coors and thus have a very small impact on Bud Light prices.) As shown, the prices observed in the post-merger periods substantially exceed the prices in the "No Merger" baseline and the price increases are almost entirely due to coordinated effects.

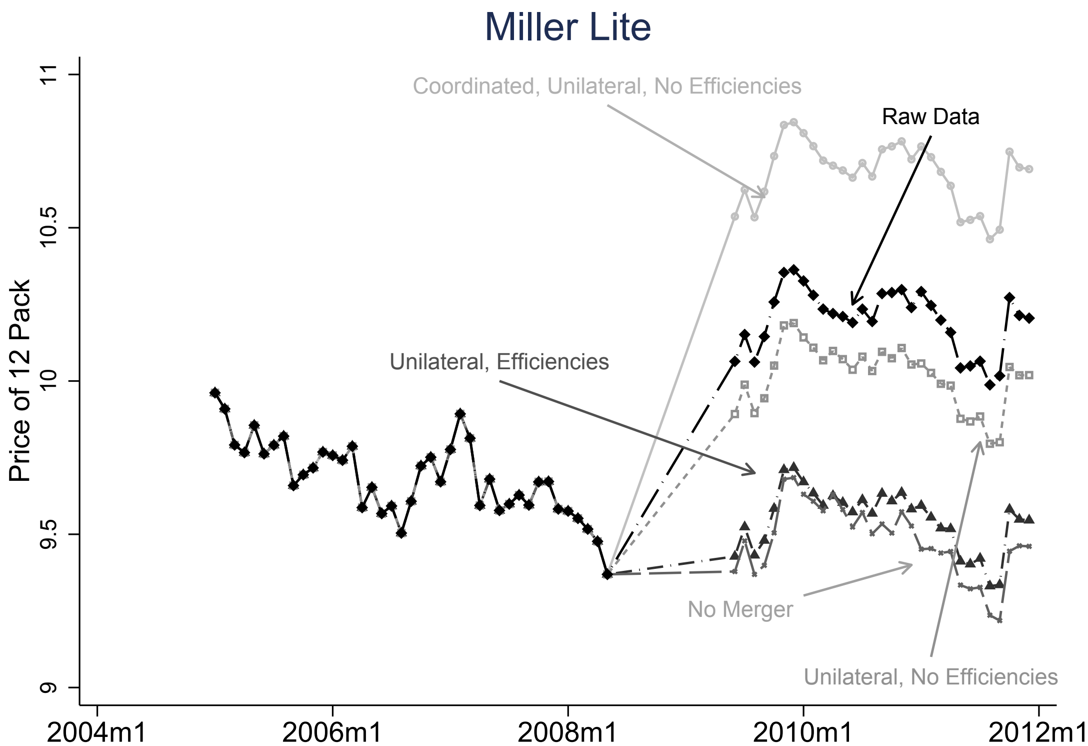

Figure 4: Counterfactuals Prices for Miller Lite

Notes: This figure plots the average retail prices of Miller Lite 12 packs in the raw data and under four different counterfactual scenarios. Each dot represents the average prices across the 39 regions.

Table X provides the mean retail prices and markups of ABI, Miller, and Coors 12 packs, along with selected welfare statistics based on all products in our sample. All numbers are for 2011, the final year of the sample. A comparison of columns (i) and (v) reveals that the merger increases ABI prices from \$9.43 in the "No Merger" scenario to \$10.03 in the raw data, while Miller prices increase from \$8.19 to \$8.94 and Coors prices increase from \$9.26 to \$10.18. The analogous comparison of markups shows smaller increases for ABI than for Miller and Coors because only the latter brands benefit from marginal cost reductions. A comparison of columns (i) and (iii) reveals that the observed prices of these firms are 6–8% higher than they would have been under Nash–Bertrand competition and the markups are 17–18% higher.

We now turn to the welfare statistics. All the numbers shown are the percentage differences relative to the "No Merger" counterfactual in which the Miller/Coors merger does not occur. Column (i) shows that the merger increases producer surplus by 22.1% relative to the no merger baseline and comparison to column (iii) reveals that a little more

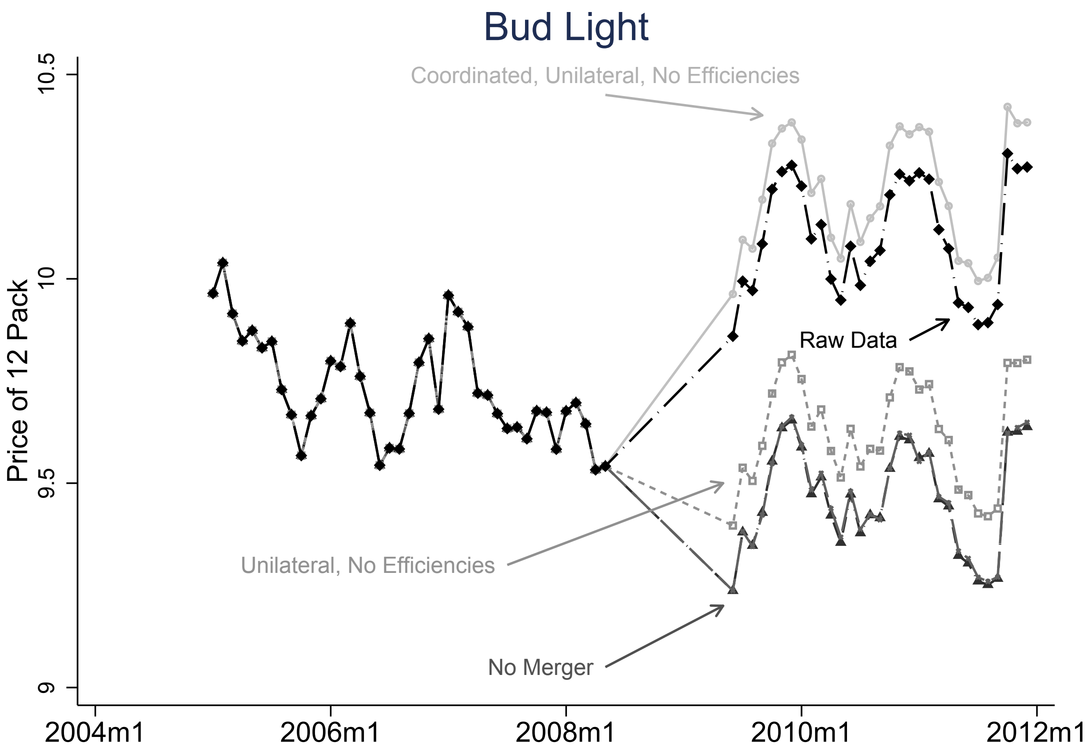

Figure 5: Counterfactuals Prices for Bud Light

Notes: This figure plots the average retail prices of Bud Light 12 packs in the raw data and under three different counterfactual scenarios. Each dot represents the average prices across the 39 regions.

than half of these gains are due to coordination. Column (i) also shows that the merger reduces consumer surplus by 3.7% relative to the no merger baseline. The consumer surplus effects vary substantially and predictably with the roles of coordination and efficiencies. For example, column (iii) shows that without coordinated effects (but with efficiencies), consumer surplus falls by only 0.2% due to the merger. This may well have been the scenario deemed most likely by the DOJ in its decision to clear the merger. Lastly, column (i) shows that the merger increases total surplus by 1.3% relative to the no merger baseline, so the antitrust clearance of the merger could be justified given a total welfare standard.

7 Conclusion

This article summarizes our empirical investigation into the economic effects of the Miller-Coors joint venture in the U.S. brewing industry. That the prices of MillerCoors and ABI increase after the Miller/Coors merger, in both absolute terms and relative to their competi-

Table X: Results from Counterfactual Analysis

Coordinated Effects:	yes	yes	no	no	no
Unilateral Effects:	yes	yes	yes	yes	no
Efficiencies:	yes	no	yes	no	no
	(i)	(ii)	(iii)	(iv)	(v)
			Retail Prices		
ABI	10.03	10.14	9.38	9.55	9.43
Miller	8.94	9.37	8.28	8.72	8.19
Coors	10.18	10.85	9.56	10.22	9.26
			Brewer Markups		
ABI	4.45	4.56	3.81	3.97	3.84
Miller	4.52	4.32	3.83	3.63	3.05
Coors	4.25	4.06	3.61	3.41	2.68
			Welfare Statistics		
Producer Surplus	22.1%	19.1%	10.3%	8.2%	·
ABI	10.3%	19.8%	-0.08%	9.3%	·
Miller	37.8%	20.2%	24.6%	9.1%	·
Coors	47.8%	12.9%	34.7%	3.5%	·
Consumer Surplus	-3.7%	-5.3%	-0.2%	-2.1%	·
Total Surplus	1.3%	-0.6%	1.8%	-0.1%	·

Notes: This table provides volume-weighted mean prices and markups, separately, for 12 pack flagship brands of ABI, Miller, and Coors, under five different economic scenarios. Also shown are the percentage changes in producer surplus, consumer surplus, and total surplus relative to the "No Merger" scenario. The welfare statistics are calculated using the complete dataset (i.e., all products). Column (i) is based on the raw data and supply-side parameter estimates. Columns (ii) to (v) show the results from counterfactual scenarios. The numbers in column (ii) are computed assuming the merger occurs with coordinated and unilateral effects but without efficiencies. The numbers in column (iii) are computed assuming the merger occurs with unilateral effects and efficiencies but no coordinated effects. The numbers in column (iv) are computed assuming the merger occurs with unilateral effects but without efficiencies or coordinated effects. Lastly, the numbers in column (v) are computed assuming that the Miller/Coors merger does not occur. All statistics are for 2011.

tors, is visually evident and confirmed with econometric analysis. The magnitude of the ABI price increase, in particular, is difficult to explain with the standard model of differentiated-products Nash–Bertrand competition. Indeed, if a parameter is added to the standard model that allows MillerCoors and ABI to partially internalize their pricing externality in the post-merger periods, then Nash–Bertrand competition is rejected.

One plausible interpretation of this result is that the merger had coordinated effects. This would reinforce the expressed view of antitrust agencies that mergers can soften the intensity of price competition between the merging firm and its remaining competitors. It is somewhat novel in the empirical literature, which has focused much more on understanding how mergers affect the unilateral pricing incentives of firms, holding the equilibrium concept fixed (e.g., Deneckere and Davidson (1985); Berry and Pakes (1993); Hausman, Leonard and Zona (1994); Werden and Froeb (1994); Nevo (2000a); Jaffe and Weyl (2013)).

Our empirical analysis does not inform why the Miller/Coors merger may have had these coordinated effects. Indeed, one challenge for future research is to understand the conditions under which consolidation either enables collusion or exacerbates the impact of collusion. That said, the U.S. brewing industry does exhibit many of the characteristics that the Merger Guidelines enumerate as contributing to the likelihood of coordinated effects. Retail prices are observable and ABI and MillerCoors may also gain visibility into wholesale prices through their interactions with wholesalers and retailers. Individual sales are small and frequent, which means that firms may be more easily deterred from undertaking competitive initiatives because the short-term gain is smaller. The inelasticity of market demand suggests large gains from coordination. The bargaining power of retailers is limited by the lack of viable private label store brands and the regulatory prohibition on slotting allowances (which makes it harder for retailers to discipline coordination by auctioning shelf space).

Some business-to-consumer markets are broadly similar to brewing along many of these dimensions, such as airlines, mobile phone services, and gasoline stations. Within this class, one possibility is that price leadership (or focal point pricing, more generally) became more feasible in the brewing industry after the Miller/Coors merger reduced heterogeneity in distribution costs among the major domestic brewers; such an effect is discussed in the report by Ivaldi et al. (2003) on coordinated effects to the European Union Directorate-General for Competition. Alternatively, the increase in concentration alone could be sufficient to facilitate coordination. That said, the generality of the results is limited. Beer is distinct among branded consumer products due to the regulatory prohibition on slotting allowances and the lack of viable private label store brands. Business-to-business markets are more likely to feature privately negotiated contracts, powerful buyers, and other characteristics that make coordination more difficult. While we interpret our results as suporting the possibility of coordinated effects, prospective analyses of mergers should be grounded in the relevant industry details.

References

Abreu, Dilip, "On the Theory of Infinitely Repeated Games with Discounting," *Econometrica*, 1988, *56* (2), 383–96.

Ascher, Bernard, "Global Beer: The Road to Monopoly," 2012. American Antitrust Institute monograph.

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

___ , ___ , and ___ , "Efficiencies Brewed: Pricing and Consolidation in U.S. Brewing," *RAND Journal of Economics*, 2015, *46* (2), 328–361.

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

Berry, Steven, "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics*, 1994, *25*, 242–262.

___ and Ariel Pakes, "Some Applications and Limitations of Recent Advances in Empirical Industrial Organization: Merger Analysis," *American Economic Review*, 1993, *83* (2), 247–252.

___ and Philip A. Haile, "Identification in Differentiated Products Markets Using Market Level Data," *Econometrica*, 2014, *82* (5), 1749–1798.

___ , James Levinsohn, and Ariel Pakes, "Automobile Prices in Market Equilibrium," *Econometrica*, July 1995, *63* (4), 847–890.

Bjornerstedt, Jonas and Frank Verboven, "Does Merger Simulation Work? A 'Natural Experiment' in the Swedish Analgesics Market," *American Economic Journal: Applied Economics*, 2015. Forthcoming.

Black, R. Michael, Gregory Crawford, Shihua Lu, and Halbert White, "A Virtual Stakes Approach to Measuring Competition in Product Markets," *Mimeo*, 2004.

Bresnahan, Timothy, "Competition and Collusion in the American Automobile Oligopoly: The 1955 Price War," *Journal of Industrial Economics*, 1987, *35* (4), 457–482.

——, "Empirical Studies of Industries with Market Power," 1989. In *Handbook of Industrial Organization*, R. Schmalensee and R. Willig (eds), Vol. 2. North-Holland, Amsterdam.

Bronnenberg, Bart J., Michael W. Kruger, and Carl F. Mela, "The IRI Marketing Data Set," *Marketing Science*, 2008, *27* (4), 745–748.

Cardell, Scott N., "Variance Components Structures for the Extreme-Value and Logistic Distributions with Applications to Models of Heterogeneity," *Econometric Theory*, 1997, *13*, 185–213.

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

Conlon, Christopher T. and Nirupama S. Rao, "The Price of Liquor is Too Damn High: The Effects of Post and Hold Pricing," 2016.

Corts, Kenneth S., "Conduct Parameters and the Measurement of Market Power," *Journal of Econometrics*, 1999, *88*, 227–225.

Cragg, John G. and Stephen G. Donald, "Testing Indentifiability and Specification in InInstrument Variable Models," *Econometric Theory*, 1993, *9*, 222–240.

Dafny, Leemore S., Mark Duggan, and Subramaniam Ramanarayanan, "Paying a Premium on Your Premium? Consolidation in the U.S. Health Insurance Industry," *American Economic Review*, 2012, *102* (2), 1161–1185.

Deneckere, Raymond and Carl Davidson, "Incentives to Form Coalitions with Bertrand Competition," *The RAND Journal of Economics*, 1985, *16*, 473–86.

Friedman, James W., "A Non-cooperative Equilibrium for Supergames," *Review of Economic Studies*, 1971, *38* (1), 1–12.

Gandhi, Amit and Jean-Francois Houde, "Measuring Substitution Patterns in Differentiated Products Industries," *Working Paper*, January 2016.

Gilbert, Richard and Hillary Greene, "Merging Innovation into Antitrust Agency Enforcement of the Clayton Act," Presented at the Eigth Annual Searle Center Conference on Antitrust Economics and Competition Policy September 2015.

Goldberg, Pinelopi K. and Rebecca Hellerstein, "A Structural Approach to Identifying the Source of Local-Currency Price Stability," *Review of Economic Studies*, 2013, *80* (1), 175–210.

Grennan, Matthew, "Price Discrimination and Bargaining: Empirical Evidence from Medical Devices," *American Economic Review*, 2013, *103* (1), 145–177.

Grigolon, Laura and Frank Verboven, "Nested Logit or Random Coefficient Logit? A Comparison of Alternative Discrete Choice Models of Product Differentiation," *Review of Economics and Statistics*, 2014, *96* (5), 916–935.

Hansen, Lars P., "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 1982, *50*, 1029–1054.

Hausman, Jerry, Gregory K. Leonard, and J. Douglas Zona, "Competitive Analysis with Differentiated Products," *Annales D'Economie et de Statistique*, 1994, *34* (1), 159–180.

Hellerstein, Rebecca, "Who Bears the Cost of a Change in the Exchage Rate? Pass-Through Accounting for the Case of Beer," *Journal of International Economics*, 2008, *76*, 14–32.

Hendel, Igal and Aviv Nevo, "Sales and Consumer Inventory," *The RAND Journal of Economics*, 2006, *37* (3), 543–561.

Houde, Jean-Francois, "Spatial Differentiation and Vertical Mergers in Retail Markets for Gasoline," *American Economic Review*, 2012, *102* (5), 2147–82.

Igami, Mitsuru, "Market Power in International Commodity Trade: The Case of Coffee," *Journal of Industrial Economics*, 2015, *63* (2), 225–248.

Ivaldi, Marc, Bruno Jullien, Patrick Rey, Paul Seabright, and Jean Tirole, "The Economics of Tacit Collusion," March 2003. Final Report for DG Competition, European Commission.

Jaffe, Sonia and E. Glen Weyl, "The First Order Approach to Merger Analysis," *American Economic Journal: Microeconomics*, 2013, *5* (4), 188–213.

McClain, Joe, "2011-2012 Annual Report," Technical Report, Beer Institute, Chicago Illinois 2012.

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

Michel, Christian, "Identification and Estimation of Intra-Firm and Industry Competition via Ownership Change," 2016.

Miller, Nathan H., Marc Remer, Conor Ryan, and Gloria Sheu, "UPP as a Predictor of Merger Price Effects," *International Journal of Industrial Organization*, January 2017. Forthcoming.

Nevo, Aviv, "Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry," *The RAND Journal of Economics*, 2000, *31* (3), pp. 395–421.

___ , "A Practitioner's Guide to Estimation of Random Coefficients Logit Models of Demand," *Journal of Economics and Management Strategy*, 2000, *9*, 513–548.

___ , "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, 2001, *69* (2), pp. 307–342.

Peters, Craig, "Evaluating the Performance of Merger Simulation: Evidence from the U.S. Airline Industry," *Journal of Law & Economics*, October 2006, *49* (2), 627–49.

Pinske, Joris and Margaret E. Slade, "Mergers, Brand Competition, and the Price of a Pint," *European Economic Review*, 2004, *48*, 617–643.

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

Romeo, Charles, "Filling Out the Instrument Set in Mixed Logit Demand Systems for Aggregate Data," 2014.

___ , "Incorporating Prior Information into a GMM Objective for Mixed Logit Demand Systems," *Journal of Industrial Economics*, 2016, *64*, 336–363.

Rotemberg, Julio J. and Garth Saloner, "A Supergame-Theoretical Model of Price Wars during Booms," *American Economic Review*, 1986, *76* (3), 390–407.

Shapiro, Carl, "The 2010 Horizontal Merger Guidelines: From Hedgehog to Fox in Forty Years," *Antitrust Law Journal*, 2010, *77*, 49–107.

Slade, Margaret E., "Market Power and Joint Dominance in U.K. Brewing," *Journal of Industrial Economics*, 2004, *52* (1), 132–163.

Stock, James A. and Motohiro Yogo, "Testing for Weak Instruments in Linear IV Regression," 2005. In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, D.W.K. Andrews and J.H. Stock (eds). New York: Cambridge University Press.

Sullivan, Christopher J., "The Ice Cream Split: Empirically Distinguishing Price and Product Space Collusion," *mimeo*, 2016.

Sweeting, Andrew and Xuezhen Tao, "Dynamic Oligopoly Pricing with Asymmetric Information: Implications for Mergers," 2016.

Tremblay, Victor J. and Carol Horton Tremblay, *The U.S. Brewing Industry*, The MIT Press, 2005.

Weinberg, Matthew C. and Daniel Hosken, "Evidence on the Accuracy of Merger Simulations," *Review of Economics and Statistics*, 2013, *95* (5), 1584–1600.

Werden, Gregory J. and Luke Froeb, "The Effects of Mergers in Differentiated Products Industries: Logit Demand and Merger Policy," *Journal of Law, Economics, and Organization*, 1994, *10*, 407–426.

Wooldridge, Jeffrey M., *Econometric Analysis of Cross Section and Panel Data*, 2nd ed., MIT Press, 2010.

Wright, Jonathan, "Detecting Lack of Identification in GMM," *Econometric Theory*, 2003, *68* (6), 1055–1096.

Appendix

A Data

Table A.1 provides information on the prices and revenue shares for major beer brands based on the six months from January to June 2008. The brands are listed in the order of their revenue share. Our regression samples 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. These included brands account for 68% of all unit sales of SAB Miller, Molson Coors, ABI, Modelo, and Heineken. The most popular brands that we omit are regional brands (e.g., Yuengling Lager and 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. We also exclude some brands that enter or exit during the sample period (e.g., Budweiser Select, Bud Light Lime). While brands in this last category could be incorporated, it would require brand-specific modifications to the demand system.

We restrict our attention to 6 packs, 12 packs, and 24/30 packs. These sizes account for 75% of all unit sales among the brands that we consider. Table A.1 also provides the distribution of sales volume across these size categories for each brand listed. For example, 11% of Bud Lite is sold as 6 packs, 34% is sold as 12 packs, and 55% is sold as 24/30 packs. Because these numbers are weighted by volume, it can also be determined that more 12 packs are sold than 24/30 packs (on a unit basis). Domestic beers tend to be sold mostly as 12 packs and 24 packs, while imports tend to be sold mostly as 6 packs and 12 packs. This amplifies the average price differences shown because smaller package sizes tend to be more expensive on a per-volume basis.

We 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.[18] Table A.2 provides the region-specific HHI in 2011, as well as the pre-merger predicted change in HHI (ΔHHI) as of January to May 2008. There is a fair amount of cross-sectional variation in concentration. Of the

[18]The regions included in our 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.

Table A.1: Average Prices and Revenue Shares: January to June 2008

Brand	Price	Revenue Share	Fraction of Sales: 6 Pack	12 Pack	24/30 Packs
Bud Light	9.45	12.2	10.7	34.0	55.3
Miller Lite	9.42	7.8	8.0	31.1	61.0
Coors Light	9.47	6.0	10.2	33.5	56.3
Budweiser	9.46	5.9	14.4	34.0	60.4
Corona Extra	14.54	5.7	17.7	66.8	15.5
Heineken	14.65	3.3	24.9	70.8	4.3
Busch Light	6.95	3.1	2.1	23.5	74.5
Natural Light	6.48	2.9	5.9	31.5	62.6
Yuengling Lager	9.61	2.7	19.0	60.2	20.8
Corona Light	14.70	2.1	24.4	72.1	3.5
Michelob Ultra	10.91	2.1	24.5	72.5	3.0
Miller High Life	7.23	2.0	7.1	46.8	46.1
Busch	7.00	1.8	3.6	31.9	64.5
Miller Genuine Draft	9.45	1.4	15.5	39.7	44.8
Michelob Light	10.84	1.3	25.8	73.2	1.0
Labatt Blue	9.23	1.3	1.9	35.6	62.5
Keystone Light	6.33	1.2	0.3	19.9	79.8
Blue Moon	14.65	1.1	47.5	52.5	0.0
Budweiser Select	9.47	1.0	12.2	46.7	45.2
Heineken Light	14.87	1.0	24.1	74.1	1.8
Natural Ice	6.45	1.0	6.9	47.1	46.1
Pabst Blue Ribbon	6.99	0.8	3.4	54.2	42.4
Tecate	11.51	0.8	9.3	39.3	51.4
Modelo Especial	14.25	0.7	21.4	76.6	2.1
Coors	9.47	0.6	6.0	39.1	54.5
Bud Light Lime	12.93	0.5	46.8	53.2	0.0

Notes: This table provides summary statistics on the major beer brands. Price is the ratio of revenue to 144 oz-equivalent unit sales. Revenue share is the total revenue of the brand divided by total revenue in the beer category. The remaining three columns show the fraction of revenues derived from six, 12, and 24/30 packs, respectively. The calculations are based on the IRI supermarket data from January through May 2008.

39 regions, 23 have post-merger HHIs that are above the threshold of 2,500 that the Merger Guidelines recognize as delineating "highly concentrated" markets.

McClain (2012) reports that supermarkets account for 20% of off-premise beer sales. The other major sources of off-premise beer sales are liquor stores (38%), convenience stores (26%), mass retailers (6%), and drugstores (3%). The IRI Academic Database includes information on sales in drugstores. In the next appendix section, we show that retail price patterns in that channel are similar to those in supermarkets. We do not have data for the other channels.

Table A.2: HHIs and Predicted Changes in HHI by IRI Region

Region	HHI	ΔHHI	Region	HHI	ΔHHI
Atlanta	2,120	367	Birmingham/Montgomery	2,989	400
Boston	1,925	188	Buffalo/Rochester	1,439	376
Charlotte	2,867	436	Chicago	2,618	484
Cleveland	1,815	400	Dallas	2,860	715
Des Moines	3,171	275	Detroit	2,372	311
Grand Rapids	2,864	311	Green Bay	3,537	448
Hartford	2,717	220	Houston	2,602	295
Indianapolis	3,382	1,022	Knoxville	3,009	371
Los Angeles	1,851	249	Milwaukee	3,718	472
Mississippi	3,647	417	West Texas/New Mexico	2,981	362
New Orleans	2,879	475	New York	1,792	216
Omaha	3,104	318	Peoria/Springfield	3,077	555
Phoenix	2,625	424	Portland, OR	1,551	479
Raleigh/Durham	2,498	265	Richmond/Norfolk	2,599	325
Roanoke	2,929	450	Sacramento	1,672	296
San Diego	1,644	353	San Francisco	1,422	210
Seattle/Tacoma	1,558	370	South Carolina	3,413	368
Spokane	2,528	684	St Louis	3,694	143
Syracuse	1,641	313	Toledo	3,059	396
Washington DC	1,711	289			

Notes: The table shows the post-merger HHIs calculated as the sum of squared market shares in 2011 and the pre-merger predicted change in HHIs (ΔHHI) based on market shares in the first five months of 2008. The market shares are calculated based on each brewer's share of total sales in the data. The data are not restricted to the brands/sizes studied in the empirical model and the market shares do not incorporate the outside good.

B Descriptive Retail Price Regressions

This section addresses questions that may arise about the descriptive regressions in Section 3, related to the store-level composition of the IRI data, the impact of promotions, and whether the results extend beyond the supermarket channel. We apply the differences-in-differences specification shown in equation (14) to store-level data, replacing product×region fixed effects with product×store fixed effects. For brevity, we consider specification with product-specific trends and no other controls. The dependent variables include the average price, the frequency of promotions, the regular price, and the promotion price. Promotions are not observed, but we follow Hendel and Nevo (2006) and define a price as being promotional if it is less than 50% of the highest price in the preceding month.

Table B.1 provides the results. The basic results that we document in Section 3 hold if store–week data are used rather than region–month data (column 1). There is some evidence that promotions are less frequent after the merger, although this effect is less pronounced

Table B.1: Supplementary Descriptive Price Regressions

Dependent Variable:	Average Price	Promotion Indicator	Regular Price	Promotion Price	Average Price
Unit of Geography:	Store	Store	Store	Store	Region
Periodicity:	Weekly	Weekly	Weekly	Weekly	Monthly
Sector:	Supermarket	Supermarket	Supermarket	Supermarket	Drugstores
$\mathbb{1}$\{MillerCoors\}	0.047	0.013	0.056	0.051	0.042
$\times\mathbb{1}$\{Post-Merger\}	(0.004)	(0.009)	(0.005)	(0.005)	(0.007)
$\mathbb{1}$\{ABI\}	0.038	0.019	0.046	0.038	0.042
$\times\mathbb{1}$\{Post-Merger\}	(0.005)	(0.011)	(0.005)	(0.004)	(0.006)
$\mathbb{1}$\{Post-Merger\}	-0.008	-0.037	-0.017	-0.020	-0.005
	(0.003)	(0.009)	(0.004)	(0.004)	(0.004)
Observations	15,408,503	15,408,503	12,085,773	3,322,730	100,587

Notes: We use OLS for estimation. The observations in the first four columns are at the brand–size–store–week–year level. The observations in the final column are at the brand–size–region–month–year level. All regressions include product (brand×size) fixed effects interacted with store fixed effects, as well as product-specific linear time trends. Standard errors are clustered at the regional level and shown in parentheses.

for MillerCoors and ABI (column 2). The regular and promotional prices seem to change in similar ways over the sample period (column 3 and 4). Taken together, the results indicate that the most important effect of the merger is on the overall price level, rather than on the frequency or magnitude of promotions. Finally, similar average price results are obtained from the drugstore sector (column 5). This can be seen graphically in Figure B.1. Average prices are more volatile due to relatively thinner sales, but the same empirical patterns are apparent. Ideally, we would also be able to verify that prices increased at convenience stores and liquor stores as well, but we were unable to obtain scanner data for these retailers. However, we would be surprised if wholesale prices increased very differently across retailers within a region, because they are legally required to buy from the same distributors.

C Numerical analysis

We provide two numerical exercises in which we perturb the estimated demand derivatives and examine the implications for estimates of the κ parameter. First, we assess the extent to which the estimates of κ could be overstated if the baseline RCNL specification does too little to relax the independence of irrelevant alternatives property of the logit demand system. To clarify this potential source of bias, consider that consumer heterogeneity likely results in some consumers who prefer domestic beer and others who prefer imported beer.

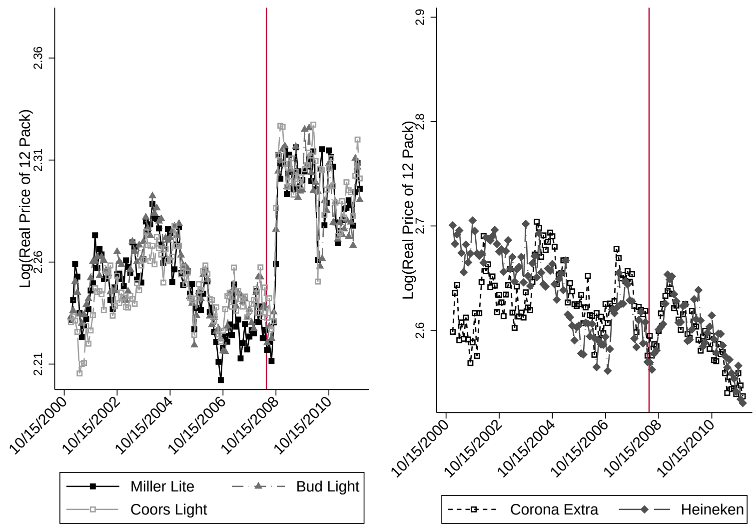

Figure B.1: Average Retail Prices of Flagship Brand 12 Packs: Drugstores

If this heterogeneity is not fully captured in the demand model, then substitution between domestic beer and imports would be overstated and substitution among domestic beers (e.g., between ABI and MillerCoors) would be understated. The supply-side implication is that the model would then understate the extent to which ABI's prices would increase with the Miller/Coors merger in Nash–Bertrand equilibrium. Because the κ parameter is identified based on whether observed ABI prices increase by more than what is predicted in Nash–Bertrand equilibrium, understating consumer heterogeneity in tastes for imports/exports would thus cause estimates of κ to be too large.

The numerical exercise involves scaling down the estimated demand derivatives between domestic beers and imports (and vice versa) according to some amount $\phi \in [0,1]$. The lost substitution is reassigned to competing brands of the same type. If $\phi = 1$, this produces markets in which there is zero substitution between domestic beers and imports and, if $\phi = 0$, then the estimated demand derivatives are unaffected. Regardless of ϕ, there is no effect on the diagonal of the demand derivatives matrix, so the own price elasticities are unchanged and substitution with the outside good is also unchanged.

To be clear about the mathematics of the exercise, we reestimate the supply side of the model plugging a scaled derivative matrix $\widetilde{\frac{\partial s_t}{\partial p_t}} = \widehat{\frac{\partial s_t}{\partial p_t}} + \Delta(\phi)$ into equation (12), where $\widehat{\frac{\partial s_t}{\partial p_t}}$ is the matrix of estimated derivatives and $\Delta(\phi)$ contains the adjustments. Consider a region–period combination with five products: Bud Light, Coors Light, Miller Lite, Corona Extra, and Heineken, respectively. Let the elements of the estimated demand derivative matrix be

$$\widehat{\frac{\partial s_t}{\partial p_t}} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{bmatrix}$$

where $a_{21} = \partial s_{2t}/\partial p_{1t}$. The adjustment matrix is given by

$$\Delta(\phi) = \begin{bmatrix} 0 & \frac{\phi(a_{42}+a_{52})a_{12}}{a_{12}+a_{32}} & \frac{\phi(a_{43}+a_{53})a_{13}}{a_{13}+a_{23}} & -\phi a_{14} & \phi a_{15} \\ \frac{\phi(a_{41}+a_{51})a_{21}}{a_{21}+a_{31}} & 0 & \frac{\phi(a_{43}+a_{53})a_{23}}{a_{13}+a_{23}} & -\phi a_{24} & -\phi a_{25} \\ \frac{\phi(a_{41}+a_{51})a_{31}}{a_{21}+a_{31}} & \frac{\phi(a_{42}+a_{52})a_{32}}{a_{12}+a_{32}} & 0 & -\phi a_{34} & -\phi a_{35} \\ -\phi a_{41} & -\phi a_{42} & -\phi a_{43} & 0 & \phi(a_{15}+a_{25}+a_{35}) \\ -\phi a_{51} & -\phi a_{52} & -\phi a_{53} & \phi(a_{14}+a_{24}+a_{34}) & 0 \end{bmatrix}$$

The first column contains adjustments to the share derivatives with respect to the Bud Light price. The diagonal element is zero, ensuring that the own price derivative (and elasticity) is unaffected. The fourth and fifth elements show reduced substitution to Corona Extra and Heineken. The total lost substitution is $\phi(a_{41}+a_{51})$ and this is reassigned to Coors Light and Miller Lite. Some assumption on the allocation between these domestic brands is required and we weight by the magnitude of the estimated substitution. This explains the second and third elements. The other columns are analogous. Each of the columns sums to zero, so that substitution with the outside good is unaffected.

Table C.1 provides the results of the first numerical exercise. We show results generated with the estimated demand derivatives of RCNL-1 (panel A) and RCNL-3 (panel B) and $\phi = 1.00, 0.80, \ldots, 0.20$. In each case, the estimate of κ is diminished relative to the baseline estimates of 0.241 (RCNL-1) and 0.291 (RCNL-3). This is because $\phi > 0$ results in greater substitution between ABI and MillerCoors and thus a greater price increase for ABI due to unilateral effects (i.e., in Nash–Bertrand equilibrium). If substitution between domestic and import brands is completely eliminated ($\phi = 1.00$), the κ estimates are reduced to 0.176 and 0.206, respectively. Thus, to the extent that the baseline demand specification does not fully capture consumer heterogeneity in tastes for imports, this explains at most 36% of the baseline κ estimate.

In our second exercise, we scale the entire estimated demand derivative matrix by a single constant, ψ, that we normalize at different levels ($\psi = 0.70, 0.80, ..., 1.20, 1.30$). This

Table C.1: Supply-Side Estimates with Adjusted Demand Derivatives (1)

	Panel A: RCNL-1 Specification				
	$\phi = 1.00$	$\phi = 0.80$	$\phi = 0.60$	$\phi = 0.40$	$\phi = 0.20$
Post-Merger Internalization of Coalition Pricing	0.176	0.197	0.214	0.231	0.247
	(0.024)	(0.024)	(0.024)	(0.024)	(0.025)
	Panel B: RCNL-3 Specification				
	$\phi = 1.00$	$\phi = 0.80$	$\phi = 0.60$	$\phi = 0.40$	$\phi = 0.20$
Post-Merger Internalization of Coalition Pricing	0.206	0.223	0.239	0.254	0.270
	(0.019)	(0.019)	(0.020)	(0.021)	(0.022)

Notes: This table shows the supply-side results obtained with RCNL demand derivative matrices that are adjusted by $\phi = 1.00, 0.80, 0.60, 0.40, 0.20$ prior to supply-side estimation. This reallocates substitution between domestic and import brands to substitution among brands of the same type; substitution across types is eliminated if $\phi = 0$. There are 94,656 observations at the brand–size–region–month–year level. All regressions incorporate a marginal cost function with the baseline marginal cost shifters and fixed effects. Standard errors are clustered by region and shown in parentheses. The standard errors are not adjusted to account for the incorporation of demand-side estimates.

approach makes demand less elastic if $\psi < 1$ and more elastic if $\psi > 1$. Adapting the brewer first-order conditions shows that this is equivalent to multiplying brewer markups by $1/\psi$:

$$p_t = mc_t - \left[\Omega_t(\kappa) \circ \left(\psi \frac{\partial s_t(p_t;\theta)}{\partial p_t}\right)^T\right]^{-1} s_t(p_t;\theta) \tag{C.1}$$

$$= mc_t - (1/\psi)\left[\Omega_t(\kappa) \circ \left(\frac{\partial s_t(p_t;\theta)}{\partial p_t}\right)^T\right]^{-1} s_t(p_t;\theta) \tag{C.2}$$

The numerical adjustment does not affect relative substitution patterns between products (including the outside good). Diversion is unchanged. However, the adjustment does allow us to investigate how supply-side inferences are affected by the overall demand elasticity. We estimate the supply side with the same methods; the demand derivatives from column (i) of Table IV are simply adjusted before incorporation into equation (12).

Table C.2 provides the results of the second exercise based on the derivatives of RCNL-1 (panel A) and RCNL-3 (panel B). We obtain smaller estimates of κ if demand is less elastic (i.e., if brewer markups are larger) and larger estimates of κ if demand is more elastic. The estimates range from 0.152 and 0.216 ($\psi = 0.70$) to 0.225 and 0.357 ($\psi = 1.30$). The null of post-merger Nash–Bertrand pricing is rejected in each instance. Thus, our main econometric finding is robust across a range of elasticities centered around the baseline point estimates. Alternative specifications of demand that result in higher or lower elasticities but which do not affect relative substitution patterns should not be expected to change the main results.

Table C.2: Supply-Side Estimates with Adjusted Demand Derivatives (2)

	Panel A: RCNL-1 Specification					
	$\psi = 0.70$	$\psi = 0.80$	$\psi = 0.90$	$\psi = 1.10$	$\psi = 1.20$	$\psi = 1.30$
Post-Merger Internalization of Coalition Pricing	0.183 (0.022)	0.211 (0.023)	0.238 (0.025)	0.289 (0.027)	0.313 (0.028)	0.336 (0.028)
	Panel B: RCNL-3 Specification					
	$\psi = 0.70$	$\psi = 0.80$	$\psi = 0.90$	$\psi = 1.10$	$\psi = 1.20$	$\psi = 1.30$
Post-Merger Internalization of Coalition Pricing	0.212 (0.020)	0.238 (0.021)	0.262 (0.022)	0.309 (0.024)	0.331 (0.025)	0.352 (0.026)

Notes: This table shows the supply-side results obtained with RCNL demand derivative matrices that are multiplied/scaled by an amount $\psi = 0.70, 0.80, ..., 1.30$. This approach dampens or amplifies the magnitude of substitution but maintains the relative substitution patterns. There are 94,656 observations at the brand–size–region–month–year level. All the regressions incorporate a marginal cost function with the baseline marginal cost shifters and fixed effects. Standard errors are clustered by region and shown in parentheses. The standard errors are not adjusted to account for the incorporation of demand-side estimates.

D Estimation details

D.1 Power of the demand-side instruments

In this section, we evaluate the relevance of excluded instruments in the RCNL-1 and RCNL-3 specifications shown in the baseline demand results using the approach of Gandhi and Houde (2016). Our GMM estimates, $\hat{\theta}$, are obtained by minimizing the objective function $\omega(\theta)'ZA^{-1}Z'\omega(\theta)$. It is possible to derive a corresponding Gauss–Newton regression equation by linearizing the residual function $\omega(\theta)$ about the true parameter value θ^0, yielding

$$\begin{aligned} \omega_{jrt}(s_{rt};\theta) &= \sum_k (\theta_k - \theta_k^0)\frac{\partial \omega_{jrt}(s_{rt};\theta^0)}{\partial \theta_k} + \sigma_j + \sigma_t + \xi_{jrt} + e_{jrt} \\ &= J_{jrt}(s_{rt};\theta^0)b + \sigma_j + \sigma_t + \xi_{jrt} + e_{jrt} \end{aligned} \tag{D.1}$$

where $J_{jrt}(s_{jrt};\theta^0)$ is a row vector of partial derivatives with k^{th} element $\frac{\partial \omega_{jrt}(s_{rt};\theta^0)}{\partial \theta_k}$, b is a vector with k^{th} element $(\theta_k - \theta_k^0)$, and e_{jrt} contains higher-order terms in the Taylor expansion. The Jacobian terms $J_{jrt}(s_{jrt};\theta^0)$ are functions of market shares, which, in turn, depend upon the structural error term ξ_{jrt}. Thus they are not orthogonal to ξ_{jrt}. However, the demand instruments can be used to form moment conditions and, when the residual functions and Jacobian terms are evaluated at the GMM estimate $\hat{\theta}$, the linear GMM estimate of equation (D.1) with weight matrix A^{-1} is $\hat{b} = 0$.

Tests for weak identification can be constructed by computing standard instrument relevance diagnostics from the firs-stage equations corresponding to equation (D.1):

$$\frac{\partial \omega_{jrt}(s_{rt}, \theta)}{\partial \theta_k} = \sigma_j + \sigma_t + \pi_k z_{jrt} + u_{jrt,k} \tag{D.2}$$

In the specific case in which θ_k is the price coefficient, α, the dependent variable, is price (i.e., $\frac{\partial \omega_{jrt}(\theta)}{\partial \alpha} = p_{jrt}$). This can be ascertained from equation (5) and motivates the first-stage regressions shown in many random coefficient logit applications (e.g., Nevo (2001)). For the other demand parameters, the dependent variables in these regressions must be obtained numerically. We use symmetric two-sided finite differences to obtain approximations (perturbations of 1e-10). Complications arise because the Jacobian is evaluated at parameter estimates rather than the population parameters, but Wright (2003) shows that Cragg and Donald (1993) tests based on the rank of the first-stage matrix will be conservative, in the sense that they do not reject the null of underidentification frequently enough.[19]

Table D.1 reports the Cragg–Donald test along with heteroskedasticity-adjusted Kleibergen–Paap tests, partial F-statistics and the Angrist–Pischke F-statistics that account for multiple endogenous regressors. The Cragg–Donald statistic is high enough to reject at the 0.05 level the null hypothesis that the bias in the point estimates is greater than 10% of the nonlinear least squares bias, following the testing procedure of Stock and Yogo (2005). Most of the F-statistics well exceed the rule-of-thumb level of 10 commonly used for linear instrumental variable regressions. The exception is the F-statistic for the nesting parameter. However, robustness checks indicate that the results are not overly sensitive to scaling the market sizes or restricting the nesting parameter to specific values rather than estimating it.

D.2 Computation

Our code is written in Matlab and largely tracks that of Nevo (2000b). The main differences relate to the contraction mapping. Grigolon and Verboven (2014) show that the standard algorithm needs to be slightly adjusted to meet the conditions for a contraction mapping if the nesting parameter ρ is sufficiently large. We solve for the mean utility levels, δ_{rt}, in region r and period t by iterating over $i = 1, 2, \ldots,$ as follows:

$$\delta_{rt}^{i+1} = \delta_{rt}^{i} + (1-\rho)\ln(s_{rt}) - (1-\rho)\ln(s_{rt}(\delta_{rt}^{i})) \tag{D.3}$$

[19]We thank J.F. Houde for bringing this to our attention.

Table D.1: First-Stage Diagnostics for the RCNL Models

Panel A: RCNL-1 (Column (ii) of Table IV)

	$\frac{\partial\xi}{\partial\alpha}$	$\frac{\partial\xi}{\partial\Pi_1}$	$\frac{\partial\xi}{\partial\Pi_2}$	$\frac{\partial\xi}{\partial\Pi_3}$	$\frac{\partial\xi}{\partial\Pi_4}$	$\frac{\partial\xi}{\partial\Pi_5}$	$\frac{\partial\xi}{\partial\rho}$
Robust Partial F-statistic	26.78	253.28	154.95	265.42	-	-	21.92
Robust Angrist–Pischke F-statistic	26.78	242.97	126.79	125.69	-	-	5.48

Panel B: RCNL-3 (Column (iv) of Table IV)

	$\frac{\partial\xi}{\partial\alpha}$	$\frac{\partial\xi}{\partial\Pi_1}$	$\frac{\partial\xi}{\partial\Pi_2}$	$\frac{\partial\xi}{\partial\Pi_3}$	$\frac{\partial\xi}{\partial\Pi_4}$	$\frac{\partial\xi}{\partial\Pi_5}$	$\frac{\partial\xi}{\partial\rho}$
Robust Partial F-statistic	26.78	-	153.35	166.26	236.78	87.44	20.45
Robust Angrist–Pischke F-statistic	14.31	-	123.74	131.14	240.00	58.47	4.46

Notes: The F-statistics were calculated while clustering standard errors by city. For the RCNL-1 specification, the p-value for the Kleibergen–Paap test of the null of underidentification is 0.003 and the Cragg–Donald Wald F-statistic is 327.4. For the RCNL-3 specification, the p-value for the Kleibergen–Paap test of the null of underidentification is 0.009 and the Cragg–Donald Wald F-statistic is 187.24.

The presence of $(1-\rho)$ slows the speed of convergence. We compute the contraction mapping in C separately for each region–period combination, using a tolerance of 1e-14.

We took several steps to ensure that the estimator computes a global optimum. First, we used the Nelder–Mead non-derivative search algorithm, which is believed to be more robust than derivative-based methods Goldberg and Hellerstein (2013). Second, we passed the optimum computed with the simplex method to a Broyden–Fletcher–Goldfarb–Shanno search algorithm and verified that the optimum did not change. Third, we verified that in each case the Hessian of the objective function at the optimum is positive definite and well-conditioned, confirming that we found a local minimum. (It also means that the linear GMM estimate of b in equation (D.1) equals zero.) Further, we started the RCNL-3 specification using 100 randomly drawn starting values (constrained within reasonable bounds) to help confirm that the estimation procedure identifies a global minimum of the objective function.

D.3 Standard error adjustment

The supply-side model of price competition is estimated conditional on the demand parameters obtained from the RCNL model. We correct the supply-side standard errors to account for the uncertainty in our demand estimates. The correction is sketched out by Wooldridge (2010), although the specific formulation is tailored to our application. Let $E[g(z_{jrt}, \theta_0^S, \theta_0^D)] = 0$ denote a vector of supply-side moment conditions, where z_{jrt} is a vector of instruments for product j in region r at time t, θ_0^S is a vector of supply-side parameters, and θ_0^D is a dimensional vector of demand-side parameters. The first-order conditions of the supply-side GMM objective function are

$$0 = \Big[J_{\theta^S} g(z_{jrt}, \hat{\theta^S}, \hat{\theta^D})\Big]^T C \Big[g(z_{jrt}, \hat{\theta^S}, \hat{\theta^D})\Big] \tag{D.4}$$

where C is a weighting matrix for the supply-side moment conditions, $g(z_{jrt}, \hat{\theta^S}, \hat{\theta^D})$ is the sample analog of the moment orthogonality conditions, matrix of the sample analog moment conditions with respect to the supply parameters. Taking a mean value expansion of $g(z_{jrt}, \hat{\theta^S}, \hat{\theta^D})$ around θ_0^S allows us to rewrite the first-order conditions:

$$0 = \Big[J_{\theta^S} g(z_{jrt}, \hat{\theta^S}, \hat{\theta^D})\Big]^T C \Big[g(z_{jrt}, \theta_0^S, \hat{\theta^D}) + J_{\theta^S} g(z_{jrt}, \bar{\theta^S}, \hat{\theta^D})\Big(\hat{\theta^S} - \theta_0^S\Big)\Big] \tag{D.5}$$

Solving for $\hat{\theta^S} - \theta_0^S$ and scaling by the square root of the number of regions R gives the following expression for $\sqrt{R}\Big(\hat{\theta^S} - \theta_0^S\Big)$:

$$-\Bigg[\Big[J_{\theta^S} g(z_{jrt}, \hat{\theta^S}, \hat{\theta^D})\Big]^T C \Big[J_{\theta^S} g(z_{jrt}, \bar{\theta^S}, \hat{\theta^D})\Big]\Bigg]^{-1} \Bigg[\Big[J_{\theta^S} g(z_{jrt}, \hat{\theta^S}, \hat{\theta^D})\Big]^T C\Bigg] \sqrt{R} g(z_{jrt}, \theta_0^S, \hat{\theta^D}) \tag{D.6}$$

Now take a mean value expansion of $g(z_{jrt}, \theta_0^S, \hat{\theta^D})$ about θ_0^D.

$$g(z_{jrt}, \theta_0^S, \hat{\theta^D}) = g(z_{jrt}, \theta_0^S, \theta_0^D) + J_{\theta^D} g(z_{jrt}, \theta_0^S, \bar{\theta^D})(\hat{\theta^D} - \theta_0^D) \tag{D.7}$$

where $J_{\theta^D} g(z_{jrt}, \hat{\theta^S}, \bar{\theta^D})$ is the Jacobian matrix of the sample analog moment conditions with respect to the demand-side parameters. The term $(\hat{\theta_0^D} - \theta_0^D)$ can be rewritten in terms of the sample analog of the demand-side moment conditions and the Jacobian of the demand-side moment conditions:

$$\Big(\hat{\theta_0^D} - \theta_0^D\Big) = -\Bigg[\Big[J_{\theta^D} h(z_{jrt}^D, \hat{\theta^D})\Big]^T A \Big[J_{\theta^D} h(z_{jrt}^D, \bar{\theta^D})\Big]\Bigg]^{-1} \Bigg[\Big[J_{\theta^D} h(z_{jrt}, \hat{\theta_0^D})\Big]^T * A\Bigg] h(z_{jrt}^D, \hat{\theta_0^D}) \tag{D.8}$$

where $h(z_{jrt}^D, \hat{\theta^D})$ is the empirical analog of the vector of demand-side moment conditions and A is an estimate of the variance covariance matrix of the demand-side moment conditions. Plugging this into equation (D.6) gives a first-order representation for $\sqrt{R}(\hat{\theta^S} - \theta_0^S)$:

$$\begin{aligned} \sqrt{R}(\hat{\theta^S} - \theta_0^S) \quad = \quad & \Bigg[\Big[J_{\theta^S} g(z_{jrt}, \hat{\theta^S}, \hat{\theta^D})\Big]^T C^S \Big[J_{\theta^S} g(z_{jrt}, \bar{\theta^S}, \hat{\theta^D})\Big]\Bigg]^{-1} \Bigg[\Big[J_{\theta^S} g(z_{jrt}, \hat{\theta^S}, \hat{\theta^D})\Big]^T C\Bigg] * \\ & \sqrt{R}\Big(g(z_{jrt}, \theta_0^S, \theta_0^D) + J_{\theta^D} g(z_{jrt}, \theta_0^S, \bar{\theta^D}) * (\hat{\theta^D} - \theta_0^D)\Big) \end{aligned} \tag{D.9}$$

A consistent estimate of $Var(\hat{\theta}^S)$ is

$$\Big[G^T CG\Big]^{-1} G^T C\Omega CG \Big[G^T CG\Big]^{-1} \tag{D.10}$$

where

$$\begin{aligned} G &\equiv \Big[J_{\theta^S} g(z_{jrt}, \hat{\theta^S}, \hat{\theta^D})\Big] \\ \Omega &= \sum_{r=1}^{R} (z_r^{S'}\omega_r + F z_r^{D'}\zeta_r)(z_r^{S'}\omega_r + F z_r^{D'}\zeta_r)' \\ F &= J_{\theta^D} g(z_{jrt}, \hat{\theta}^s, \hat{\theta^D}) \Bigg[\Big[J_{\theta^D} h(z_{jrt}^D, \hat{\theta^D})\Big]^T C^D \Big[J_{\theta^D} h(z_{jrt}^D, \hat{\theta^D})\Big]\Bigg]^{-1} \Bigg[\Big[J_{\theta^D} h(z_{jrt}, \hat{\theta^D})\Big]^T C^D\Bigg] \end{aligned}$$

The Jacobians of the supply-side moments $J_{\theta^D} g(z_{jrt}, \hat{\theta}^s, \hat{\theta^D})$ and $J_{\theta^S} g(z_{jrt}, \hat{\theta}^s, \hat{\theta^D})$ were approximated by symmetric two-sided finite differences.

E Retail sector

E.1 Overview

In this appendix, we extend the supply-side model to incorporate a retail sector. The extension features linear pricing, consistent with industry regulations that prohibit slotting allowances. Brewers set their prices first; the representative retailer observes these prices and sets downstream prices accordingly. Double marginalization arises in equilibrium. The main results regarding brewer competition are largely unaffected. This is because adding retail markups reduces implied marginal costs by a commensurate amount; the presence of a retail sector is economically similar to a per-unit tax that brewers must pay. Thus, inferences about brewer markups are robust but inferences about marginal costs are not.

E.2 Model and identification

Retail prices are set by a representative retailer. Let p_t^R be a vector of retail prices during period t and let p_t^B be a vector of brewer prices. We suppress region subscripts for brevity. The retail price vector can be decomposed into the brewer prices as follows:

$$p_t^R = p_t^B + mc_t^R + markup_t^R(\lambda, p_t^B, \theta^D) \tag{E.1}$$

where λ is the retail scaling parameter and θ^D. Brewers set their prices with knowledge of equation (E.4). The resulting first-order conditions are

$$p_t^B = mc_t^B - \left[\Omega_t(\kappa) \circ \left(\frac{\partial p_t^R(p_t^B; mc_t^R, \lambda, \theta^D)}{\partial p_t^B}\right)^T \left(\frac{\partial s_t(p_t^R; \theta^D)}{\partial p_t^R}\right)^T\right]^{-1} s_t(p_t^R; \theta_d) \qquad \text{(E.2)}$$

where mc_t^B is the vector of brewer marginal costs. Note that brewer markups depend on the retail pass-through matrix $[\partial p_t^R / \partial p_t^B]$ because this determines how brewer prices affect market shares. Plugging back into the retailer pricing equation yields

$$\begin{aligned} p_t^R &= mc_t^R + mc_t^B + markup_t^R(\lambda, p_t^B, \theta^D) \\ &- \left[\Omega_t(\kappa) \circ \left(\frac{\partial p_t^R(p_t^B; mc_t^R, \lambda, \theta^D)}{\partial p_t^B}\right)^T \left(\frac{\partial s_t(p_t^R; \theta^D)}{\partial p_t^R}\right)^T\right]^{-1} s_t(p_t^R; \theta_d) \end{aligned} \qquad \text{(E.3)}$$

The marginal cost vectors are not separately identifiable, but a composite marginal cost function can be specified along the lines of equation (11). If retail markups are invariant to brewer prices, then this system is similar to the baseline supply-side model, with the distinction that implied marginal costs incorporate some unidentifiable retail markup.

An alternative approach is to assume that the representative retailer maximizes profit. The usual derivations show that the vector of product-specific retail markups is given by

$$markup_t^R(\lambda, p_t^B, \theta^D) = \lambda \left[\left(\frac{\partial s_t(p_t^R; \theta^D)}{\partial p_t^R}\right)^T\right]^{-1} s_t(p_t^R; \theta^D) \qquad \text{(E.4)}$$

This is a standard multi-product monopoly formulation with the simple tweak that $\lambda \in [0, 1]$ scales the retail markups. The retailer sells all products and internalizes the effects that the retail price of each product has on the sales of other products. For example, the retailer has both Bud Light and Miller Lite on the shelf and a lower retail price on Bud Light results in some cannibalization of Miller Lite sales. If $\lambda = 1$, then the model corresponds to the representative retailer having monopoly power over each region. If $\lambda = 0$, then the model corresponds to marginal cost pricing; this is observationally equivalent to the constant markup model. If $0 < \lambda < 1$, the model can be interpreted as corresponding to intermediate levels of retail market power, although the mapping to a fully specified model of retail oligopoly is unclear. We show how λ affects retail pass-through in the next subsection.[20]

[20] If the retail scaling parameter is to be estimated, an additional instrument is required, because retail markups are affected by unobserved costs. The obvious candidates are demand variables. The Corts cri-

E.3 Retail pass-through

Before proceeding to the results, we develop the connection between λ and retail pass-through and show how estimation can be made computationally tractable. It is useful to rewrite the retail first-order conditions as follows:

$$f(p_t^R) \equiv p_t^R - p_t^B - mc_t^R + \lambda \left[\left(\frac{\partial s_t(p_t^R;\theta^D)}{\partial p_t^R} \right)^T \right]^{-1} s_t(p_t^R;\theta^D) = 0 \qquad \text{(E.5)}$$

Following Jaffe and Weyl (2013), the retail pass-through matrix equals

$$\frac{\partial p_t^R}{\partial p_t^B} = -\left(\frac{\partial f(p_t^R)}{\partial p_t^R} \right)^{-1} \qquad \text{(E.6)}$$

The Jacobian matrix on the right-hand side depends on both the first and second derivatives of demand. Obtaining retail pass-through via numerical integration for each set of candidate supply parameters is computationally expensive. It is simpler to calculate $\partial f(p_t^R)/\partial p_t^R$ under $\lambda = 1$ and then adjust this Jacobian in accordance with the candidate λ. To clarify this procedure, we provide a closed-form expression for column n of the Jacobian:

$$\frac{\partial f^R(p^R)}{\partial p_n} = -\begin{bmatrix} 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \end{bmatrix} + \lambda \left[\frac{\partial s}{\partial p^R}^T \right]^{-1} \left[\frac{\partial^2 s}{\partial p^R \partial p_n}^T \right] \left[\frac{\partial s}{\partial p^R}^T \right]^{-1} s - \lambda \left[\frac{\partial s}{\partial p^R}^T \right]^{-1} \left[\frac{\partial s}{\partial p_n}^T \right], \qquad \text{(E.7)}$$

where the value of one in the initial vector is in the n^{th} position. In estimation, start with the Jacobian obtained under $\lambda = 1$ and then, for each vector of candidate supply-side parameters, (i) subtract the identity matrix from the initial Jacobian, (ii) scale the remainder by λ, (iii) add back the identity matrix, and (iv) take the opposite inverse to obtain a retail pass-through matrix that is fully consistent with the candidate parameter vector.

E.4 Results

Table E.1 provides the results of supply-side estimation for different normalizations of the retail scaling parameters. Two main patterns are relevant. First, the estimates of κ are

tique applies: Estimates are consistent only if retailers set markups according to equation (E.4). We forgo estimation and instead restrict λ to different levels.

Table E.1: Supply-Side Estimates with Retail Market Power

		(i)	(ii)	(iii)	(iv)	(v)	(vi)
Post-Merger Internalization of Coalition Pricing	κ	0.291 (0.047)	0.292 (0.045)	0.294 (0.045)	0.297 (0.046)	0.300 (0.047)	0.303 (0.048)
Retail Scaling Parameter	λ	0.00	0.025	0.05	0.10	0.15	0.20
Derived Statistics							
Marginal Costs < 0		0.00%	0.25%	0.49%	1.56%	4.90%	10.78%

Table E.2: Markups and Marginal Costs with Retail Market Power

	$\lambda = 0.00$	$\lambda = 0.025$	$\lambda = 0.05$	$\lambda = 0.10$	$\lambda = 0.15$	$\lambda = 0.20$
			Average Markups			
Retail	0.00	0.51	1.01	2.02	3.03	4.04
Bud Light	3.85	3.86	3.87	3.89	3.90	3.91
Coors Light	2.63	2.61	2.58	2.53	2.48	2.43
Miller Lite	3.02	3.01	2.98	2.94	2.90	2.85
			Average Composite Marginal Costs			
Bud Light	5.88	5.37	4.86	3.84	2.83	1.82
Coors Light	7.10	6.62	6.15	5.20	4.25	3.30
Miller Lite	6.66	6.18	5.70	4.74	3.78	2.82

largely unaffected by the magnitude of the retail scaling parameter. Second, the number of products for which implied marginal costs are negative increases as the retail scaling parameter increases: Under the baseline specification ($\lambda = 0$), there are no negative marginal costs but, at the highest level shown ($\lambda = 0.20$), more than 10% of the marginal costs are negative. In Table E.2, we report the average pre-merger markups and marginal costs that arise under each normalization of λ. We restrict attention to selected brand to conserve space. As shown, retail markups increase monotonically with λ. With the baseline specification ($\lambda = 0$), there are no retail markups but, at the highest level shown ($\lambda = 0.20$), the average retail markup is \$4.04. Brewer markups change little over different levels of λ, but the implied composite marginal costs decrease nearly one for one as retail markups increase.