

SUPPLEMENT TO “UNDERSTANDING THE PRICE EFFECTS OF THE
MILLERCOORS JOINT VENTURE”
(*Econometrica*, Vol. 85, No. 6, November 2017, 1763–1791)

NATHAN H. MILLER
McDonough School of Business, Georgetown University

MATTHEW C. WEINBERG
LeBow College of Business, Drexel University

APPENDIX A: DATA

TABLE SI 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 SI 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.¹ Table SII provides the region-specific HHI in 2011, as well as the pre-merger predicted change in HHI (Δ HHI) as of January to May 2008. There is a fair amount of cross-sectional variation in concentration.

Nathan H. Miller: nathan.miller@georgetown.edu

Matthew C. Weinberg: mcw325@drexel.edu

¹The 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 SI
AVERAGE PRICES AND REVENUE SHARES: JANUARY TO JUNE 2008^a

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

^aThis 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.

Of the 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) reported 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.

APPENDIX 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 (1) to store-level data, replacing product \times region fixed effects with product \times store fixed effects. For brevity, we consider specification with product-specific trends and no other controls. The dependent vari-

TABLE SII
HHIS AND PREDICTED CHANGES IN HHI BY IRI REGION^a

Region	HHI	Δ HHI	Region	HHI	Δ HHI
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			

^aThe table shows the post-merger HHIs calculated as the sum of squared market shares in 2011 and the pre-merger predicted change in HHIs (Δ HHI) 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.

ables 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 [SIII](#) 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 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 S1](#). 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.

APPENDIX 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

TABLE III
 SUPPLEMENTARY DESCRIPTIVE PRICE REGRESSIONS^a

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
$\mathbb{1}\{\text{MillerCoors}\}$	0.047	0.013	0.056	0.051	0.042
$\times \mathbb{1}\{\text{Post-Merger}\}$	(0.004)	(0.009)	(0.005)	(0.005)	(0.007)
$\mathbb{1}\{\text{ABI}\}$	0.038	0.019	0.046	0.038	0.042
$\times \mathbb{1}\{\text{Post-Merger}\}$	(0.005)	(0.011)	(0.005)	(0.004)	(0.006)
$\mathbb{1}\{\text{Post-Merger}\}$	-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

^aWe 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 \times 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.

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. 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

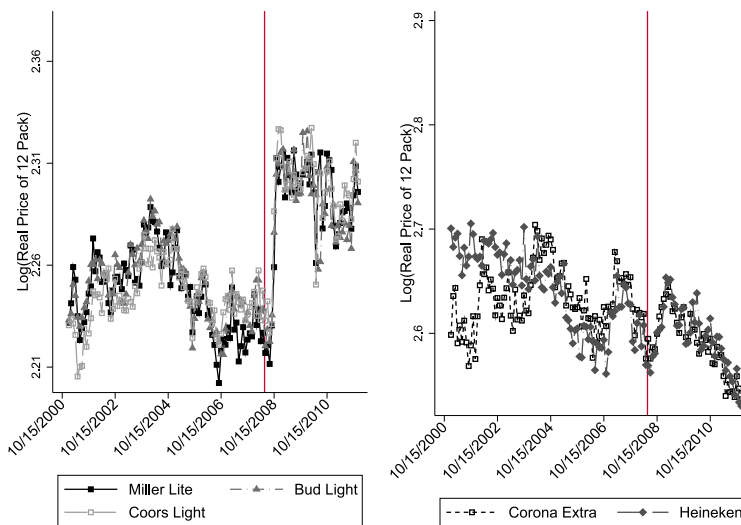


FIGURE S1.—Average retail prices of flagship brand 12 packs: drugstores.

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 SIV 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 $\phi = 1.00, 0.80, \dots, 0.20$. 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.

TABLE SIV
SUPPLY-SIDE ESTIMATES WITH ADJUSTED DEMAND DERIVATIVES (1)^a

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

^aThis 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.

In our second exercise, we scale the entire estimated demand derivative matrix by a single constant, ψ , that we normalize at different levels ($\psi = 0.70, 0.80, \dots, 1.20, 1.30$). This 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 $1/\psi$:

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

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

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 SV 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.

APPENDIX 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

TABLE SV
SUPPLY-SIDE ESTIMATES WITH ADJUSTED DEMAND DERIVATIVES (2)^a

	$\psi = 0.70$	$\psi = 0.80$	$\psi = 0.90$	$\psi = 1.10$	$\psi = 1.20$	$\psi = 1.30$
	Panel A: RCNL-1 Specification					
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					
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)

^aThis table shows the supply-side results obtained with RCNL demand derivative matrices that are multiplied/scaled by an amount $\psi = 0.70, 0.80, \dots, 1.30$. 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.

Gandhi and Houde (2016). Our GMM estimates, $\hat{\theta}$, are obtained by minimizing the objective function $\omega(\theta)' Z A^{-1} Z' \omega(\theta)$. 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

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

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 first-stage equations corresponding to equation (D.1):

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

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 (7) 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) showed that Cragg and Donald (1993) tests based on the rank of the first-stage matrix will be con-

TABLE SVI
FIRST-STAGE DIAGNOSTICS FOR THE RCNL MODELS^a

	$\frac{\partial \xi}{\partial \alpha}$	$\frac{\partial \xi}{\partial \Pi_1}$	$\frac{\partial \xi}{\partial \Pi_2}$	$\frac{\partial \xi}{\partial \Pi_3}$	$\frac{\partial \xi}{\partial \Pi_4}$	$\frac{\partial \xi}{\partial \Pi_5}$	$\frac{\partial \xi}{\partial \rho}$
Panel A: RCNL-1 (Column (ii) of Table IV)							
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)							
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

^aThe 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 under-identification 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 under-identification is 0.009 and the Cragg–Donald Wald F -statistic is 187.24.

servative, in the sense that they do not reject the null of under-identification frequently enough.²

Table SVI reports the Cragg–Donald test along with heteroscedasticity-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 (2000). The main differences relate to the contraction mapping. Grigolon and Verboven (2014) showed 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 $i = 1, 2, \dots$, as follows:

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

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

²We thank J. F. Houde for bringing this to our attention.

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 was 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

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

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:

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

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 $\sqrt{R}(\hat{\theta}^S - \theta_0^S)$:

$$\begin{aligned} & -[[J_{\theta^S} g(z_{jrt}, \hat{\theta}^S, \hat{\theta}^D)]^T C [J_{\theta^S} g(z_{jrt}, \bar{\theta}^S, \hat{\theta}^D)]]^{-1} [[J_{\theta^S} g(z_{jrt}, \hat{\theta}^S, \hat{\theta}^D)]^T C] \\ & \times \sqrt{R} g(z_{jrt}, \theta_0^S, \hat{\theta}^D). \end{aligned} \quad (\text{D.6})$$

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

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

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

$$\begin{aligned} (\hat{\theta}^D - \theta_0^D) & = -[[J_{\theta^D} h(z_{jrt}^D, \hat{\theta}^D)]^T A [J_{\theta^D} h(z_{jrt}^D, \bar{\theta}^D)]]^{-1} \\ & \times [[J_{\theta^D} h(z_{jrt}^D, \hat{\theta}^D)]^T * A] h(z_{jrt}^D, \hat{\theta}^D), \end{aligned} \quad (\text{D.8})$$

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 condi-

tions. Plugging this into equation (D.6) gives a first-order representation for $\sqrt{R}(\hat{\theta}^S - \theta_0^S)$:

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

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

$$[G^T C G]^{-1} G^T C \Omega C G [G^T C G]^{-1}, \quad (\text{D.10})$$

where

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

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.

APPENDIX 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:

$$p_t^R = p_t^B + mc_t^R + \text{markup}_t^R(\lambda, p_t^B, \theta^D), \quad (\text{E.1})$$

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

$$p_i^B = mc_i^B - \left[\Omega_i(\kappa) \circ \left(\frac{\partial p_i^R(p_i^B; mc_i^R, \lambda, \theta^D)}{\partial p_i^B} \right)^T \left(\frac{\partial s_i(p_i^R; \theta^D)}{\partial p_i^R} \right)^T \right]^{-1} s_i(p_i^R; \theta_d), \quad (\text{E.2})$$

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

$$p_i^R = mc_i^R + mc_i^B + \text{markup}_i^R(\lambda, p_i^B, \theta^D) - \left[\Omega_i(\kappa) \circ \left(\frac{\partial p_i^R(p_i^B; mc_i^R, \lambda, \theta^D)}{\partial p_i^B} \right)^T \left(\frac{\partial s_i(p_i^R; \theta^D)}{\partial p_i^R} \right)^T \right]^{-1} s_i(p_i^R; \theta_d). \quad (\text{E.3})$$

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

$$\text{markup}_i^R(\lambda, p_i^B, \theta^D) = \lambda \left[\left(\frac{\partial s_i(p_i^R; \theta^D)}{\partial p_i^R} \right)^T \right]^{-1} s_i(p_i^R; \theta^D). \quad (\text{E.4})$$

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.³

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:

$$f(p_i^R) \equiv p_i^R - p_i^B - mc_i^R + \lambda \left[\left(\frac{\partial s_i(p_i^R; \theta^D)}{\partial p_i^R} \right)^T \right]^{-1} s_i(p_i^R; \theta^D) = 0. \quad (\text{E.5})$$

³If 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 critique 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 SVII
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
<i>Derived Statistics</i>							
Marginal Costs < 0		0.00%	0.25%	0.49%	1.56%	4.90%	10.78%

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

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

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 $\partial f(p_i^R)/\partial p_i^R$ 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:

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

where the value of 1 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 SVII 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 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 SVIII, we report the average pre-merger markups and marginal costs that arise under each normalization of λ . We restrict attention to selected brands 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.

TABLE SVIII
MARKUPS AND MARGINAL COSTS WITH RETAIL MARKET POWER

	$\lambda = 0.00$	$\lambda = 0.025$	$\lambda = 0.05$	$\lambda = 0.10$	$\lambda = 0.15$	$\lambda = 0.20$
	<i>Average Markups</i>					
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
	<i>Average Composite Marginal Costs</i>					
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

REFERENCES

- CRAGG, J. G., AND S. G. DONALD (1993): "Testing Identifiability and Specification in Instrument Variable Models," *Econometric Theory*, 9, 222–240. [7]
- GANDHI, A., AND J.-F. HOUDE (2016): "Measuring Substitution Patterns in Differentiated Products Industries," Working Paper. [7]
- GOLDBERG, P. K., AND R. HELLERSTEIN (2013): "A Structural Approach to Identifying the Source of Local-Currency Price Stability," *Review of Economics and Statistics*, 80 (1), 175–210. [8]
- GRIGOLON, L., AND F. VERBOVEN (2014): "Nested Logit or Random Coefficient Logit? A Comparison of Alternative Discrete Choice Models of Product Differentiation," *Review of Economics and Statistics*, 96 (5), 916–935. [8]
- HENDEL, I., AND A. NEVO (2006): "Sales and Consumer Inventory," *The RAND Journal of Economics*, 37 (3), 543–561. [3]
- JAFFE, S., AND E. GLEN WEYL (2013): "The First Order Approach to Merger Analysis," *American Economic Journal: Microeconomics*, 5 (4), 188–213. [12]
- MCCLAIN, J. (2012): "2011–2012 Annual Report," Technical Report, Beer Institute, Chicago Illinois. [2]
- NEVO, A. (2000): "A Practitioner's Guide to Estimation of Random Coefficients Logit Models of Demand," *Journal of Economics and Management Strategy*, 9, 513–548. [8]
- (2001): "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, 69 (2), 307–342. [7]
- STOCK, J. A., AND M. YOGO (2005): "Testing for Weak Instruments in Linear IV Regression," in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, ed. by D. W. K. Andrews and J. H. Stock. New York: Cambridge University Press. [8]
- WOOLDRIDGE, J. M. (2010): *Econometric Analysis of Cross Section and Panel Data* (second Ed.). MIT Press. [9]
- WRIGHT, J. (2003): "Detecting Lack of Identification in GMM," *Econometric Theory*, 68 (6), 1055–1096. [7]

Co-editor Liran Einav handled this manuscript.

Manuscript received 25 March, 2015; final version accepted 2 March, 2017; available online 8 June, 2017.