Online Appendix

Oligopolistic Price Leadership and Mergers: The United States Beer Industry

Nathan H. Miller*
Georgetown University

Gloria Sheu†
Federal Reserve Board

Matthew C. Weinberg‡
The Ohio State University

A Data

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

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

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

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

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

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

Miller and Weinberg (2017) restrict attention to 39 of the 47 geographic regions in the IRI academic database, dropping a handful of regions in which either few supermarkets are licensed to sell beer or supermarkets are restricted to selling low-alcohol beer.[1] The IRI data are not intended to be fully representative of regions in the United States. Bronnenberg et al. (2008, p. 746) states that

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

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

In our sample, we also exclude New Orleans and San Diego, leaving a total of 37 geographic regions in the data. The exclusion is due to a computational issue related to the demographic draws. When we compute the fringe's best responses to even very small supermarkups, such as $m = 0.01$, we find that Modelo and Heineken respond by *lowering* prices in New Orleans and San Diego, often by an order of magnitude more than the supermarkup. The magnitude and direction of this response—and note that prices are acting as strategic substitutes here—makes any positive supermarkup unprofitable. We attribute this outsize strategic substitutes response to poorly-performing demographic draws. In numerical checks based on logit demand, which omits the impact of the demographic draws, we always compute PLE that feature positive supermarkups (Appendix C.3). Consistent with this,

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

the income draws of New Orleans and San Diego are much larger on average than nearly all other regions, and also exhibit greater variance.[2] Thus, we believe the computational issue is due to the demand-side of the model and is not indicative of a broader supply-side performance issue.

B Theoretical Matters

B.1 Slack Functions and the Timing Parameter

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

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

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

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

If we set η such that

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

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

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

Furthermore, if $\tau_1 = 1$, then $\psi = 1$.

QED

B.2 The Leader's Constrained Maximization Problem

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

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

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

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

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

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

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

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

Substituting into the other first order conditions gives

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

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

C Numerical Matters

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

C.1 Imputation with the Baseline Model

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

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

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

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

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

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

1. Set $\widetilde{\Delta m}^{(0)} = \overrightarrow{0}$. Search for the $\widetilde{m}_R$ that solves (C.1). We use fminsearch in Matlab with a tolerance of 1e-4.[3] Each evaluation of the loss function requires a number of numerical steps, so the search is nontrivial. However, it is feasible nonetheless because there is only one unknown. At the solution, the IC constraint binds, but the dispersion of supermarkups across regions is sub-optimal. Denote the solution $\widetilde{m}_R^{(0)}\left(\widetilde{\Delta m}^{(0)}\right)$.

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

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

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

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

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

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

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

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

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

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

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

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

6. Repeat step 3.

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

C.2 Imputation with Alternative Assumptions

Size-Specific Supermarkups

We now consider the case in which the leader sets two supermarkups: one that applies to 6 and 12 packs, and another that applies to 24 packs. There are $2R$ supermarkups that must be recovered. Denote the supermarkups for the 6 and 12 packs as $m^1 = (m^1_1, m^1_2, \ldots, m^1_R)$ and the markups for the 24 packs as $m^2 = (m^2_1, m^2_2, \ldots, m^2_R)$. An evaluation of the leader's Lagrangian provides the following first order conditions:

$$\begin{aligned}
g_k(m^1, m^2; \eta) &\equiv \sum_{r=1}^{R} \left(\frac{1}{1-\eta} \pi^{PL}_{kr}(m^1_r, m^2_r) - \pi^{D,k}_{kr}(m^1_r, m^2_r) - \frac{\eta}{1-\eta} \pi^{B}_{kr} \right) = 0 \\
h^1_r(m^1_r, m^1_R, m^2_r, m^2_R) &\equiv \frac{\frac{\partial \pi^{PL}_{1r}(m^1_r, m^2_r)}{\partial m^1_r}}{\frac{\partial \pi^{D,k}_{kr}(m^1_r, m^2_r)}{\partial m^1_r} - \frac{\partial \pi^{PL}_{kr}(m^1_r, m^2_r)}{\partial m^1_r}} - \frac{\frac{\partial \pi^{PL}_{1R}(m^1_R, m^2_R)}{\partial m^1_R}}{\frac{\partial \pi^{D,k}_{kR}(m^1_R, m^2_R)}{\partial m^1_{Rt}} - \frac{\partial \pi^{PL}_{kR}(m^1_R, m^2_R)}{\partial m^1_{Rt}}} = 0 \\
h^2_r(m^1_r, m^1_R, m^2_r, m^2_R) &\equiv \frac{\frac{\partial \pi^{PL}_{1r}(m^1_r, m^2_r)}{\partial m^2_r}}{\frac{\partial \pi^{D,k}_{kr}(m^1_r, m^2_r)}{\partial m^2_r} - \frac{\partial \pi^{PL}_{kr}(m^1_r, m^2_r)}{\partial m^2_r}} - \frac{\frac{\partial \pi^{PL}_{1R}(m^1_R, m^2_R)}{\partial m^1_R}}{\frac{\partial \pi^{D,k}_{kR}(m^1_R, m^2_R)}{\partial m^1_{Rt}} - \frac{\partial \pi^{PL}_{kR}(m^1_R, m^2_R)}{\partial m^1_{Rt}}} = 0
\end{aligned}$$

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

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

It is theoretically possible that a solution to the set of $2R$ equations provided above (i.e. an interior solution) does not exist. This could occur if searching for a solution under the assumption that i is the binding firm generates $g_j(m^1, m^2; \eta) < 0$ (i.e., that the IC constraint is violated for firm j), but searching for a solution using j as the binding firm generates $g_i(m^1, m^2; \eta) < 0$.[4] Restated loosely, one coalition firm might prefer higher supermarkups for smaller package sizes, and another coalition firm might prefer higher supermarkups for larger package sizes. This does not occur in our application because the relative shares of the coalition firms are similar for the smaller and larger package sizes.

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

Independent Regions

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

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

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

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

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

C.3 Numerical Evidence on Imputation

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

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

2. **Equilibrium Computation:** Compute the solution to the constrained maximization

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

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

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

C.4 Simulation

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

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

D Results with Alternative Assumptions

D.1 Overview

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

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

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

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

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

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

D.2 Results

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

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

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

We now turn to the alternative model with independent regions, such that IC constraints are not pooled and multi-market contact does not matter for the results. As the

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

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

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

E The Demand System

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

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

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

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

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

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

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

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

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

F Adding a Retail Sector

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

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

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

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

$$\pi_i(\mu^w) = \sum_{j \in \mathbb{J}_i} \mu_j^w q_j(mc + \mu^w + \mu^r) \tag{F.2}$$

$$= \sum_{j \in \mathbb{J}_i} \mu^w q_j(\mu^w + \phi) \tag{F.3}$$

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

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

G Additional Tables and Figures

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

Panel A: OLS Regression Results					
	Bertrand	$\eta = 0.25$	$\eta = 0.26$	$\eta = 0.30$	Unconstrained
ABI×Post-Merger	0.657	0.009	-0.031	-0.210	0.504
	(0.094)	(0.118)	(0.121)	(0.122)	(0.147)
Miller×Post-Merger	0.189	-0.480	-0.523	-0.711	-0.037
	(0.074)	(0.012)	(0.099)	(0.103)	(0.133)
Coors×Post-Merger	-0.067	-0.756	-0.798	-0.986	-0.278
	(0.095)	(0.110)	(0.110)	(0.117)	(0.169)
Distance	0.22	0.231	0.232	0.237	0.331
	(0.066)	(0.066)	(0.067)	(0.068)	(0.120)
Panel B: Other Statistics					
	Bertrand	$\eta = 0.25$	$\eta = 0.26$	$\eta = 0.30$	$\eta = 1.00$
$\overline{m}_{2006}^{6,12}$	0	1.02	1.07	1.27	3.99
$\overline{m}_{2006}^{24}$	0	1.13	1.18	1.43	5.08
$\overline{m}_{2007}^{6,12}$	0	1.07	1.12	1.33	4.11
$\overline{m}_{2007}^{24}$	0	1.18	1.24	1.49	5.18
$\overline{m}_{2010}^{6,12}$	0	1.64	1.73	2.10	4.24
$\overline{m}_{2010}^{24}$	0	1.71	1.80	2.22	4.94
$\overline{m}_{2011}^{6,12}$	0	1.72	1.81	2.20	4.54
$\overline{m}_{2011}^{24}$	0	1.75	1.84	2.27	4.97
Marginal Costs < 0	0.001	0.005	0.006	0.010	0.163

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

Table G.2: Model Selection with Independent Regions

Panel A: OLS Regression Results					
	Bertrand	$\eta = 0.25$	$\eta = 0.27$	$\eta = 0.30$	Unconstrained
ABI×Post-Merger	0.657	0.087	0.012	-0.114	0.607
	(0.094)	(0.096)	(0.097)	(0.099)	(0.147)
Miller×Post-Merger	0.189	-0.429	-0.508	-0.64	-0.017
	(0.074)	(0.068)	(0.069)	(0.070)	(0.137)
Coors×Post-Merger	-0.067	-0.698	-0.778	-0.911	-0.242
	(0.095)	(0.097)	(0.098)	(0.101)	(0.173)
Distance	0.22	0.227	0.227	0.228	0.34
	(0.066)	(0.067)	(0.067)	(0.067)	(0.120)

Panel B: Other Statistics					
	Bertrand	$\eta = 0.25$	$\eta = 0.27$	$\eta = 0.30$	$\eta = 1.00$
$\overline{m}_{2006}$	0	1.11	1.22	1.39	4.51
$\overline{m}_{2007}$	0	1.14	1.25	1.43	4.65
$\overline{m}_{2010}$	0	1.64	1.82	2.11	4.63
$\overline{m}_{2011}$	0	1.7	1.89	2.19	4.79
Marginal Costs < 0	0.001	0.005	0.006	0.009	0.17

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

References

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

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

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

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.

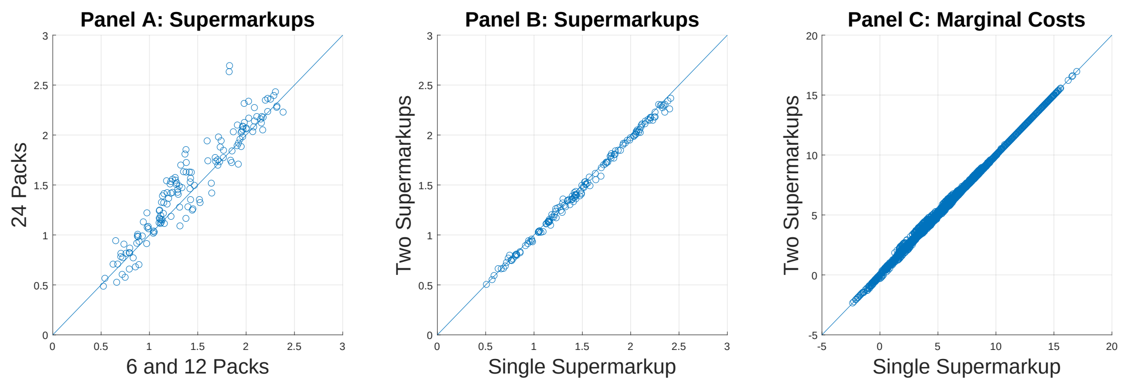

Figure G.1: Size-Specific Supermarkups and the Baseline Model

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

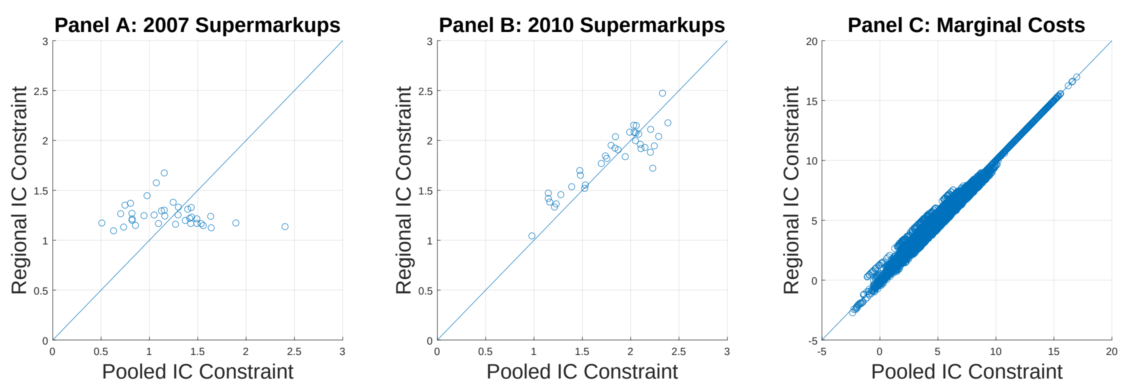

Figure G.2: Regional IC Constraints vs. the Baseline Model

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

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

Miller, N. H. and M. C. Weinberg (2017). Understanding the Price Effects of the MillerCoors

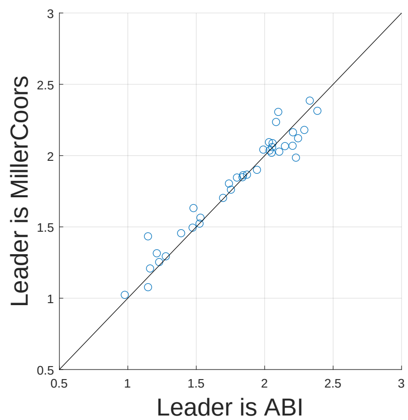

Figure G.3: Supermarkups with ABI and MillerCoors as Leader

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

Joint Venture. *Econometrica 85*(6), 1763–1791.

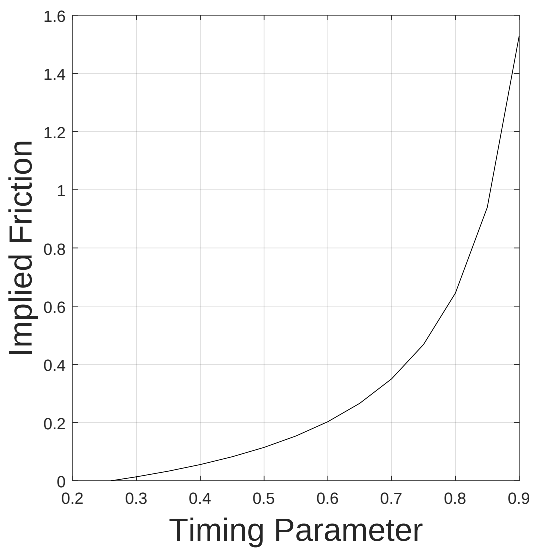

Figure G.4: Slack Functions in the Numerical Illustration

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