Online Appendix for: “The Rise of Fringe Competitors In the Wake of an Emerging Middle Class: An Empirical Analysis”

Alon Eizenberg*
Department of Economics
Hebrew University of Jerusalem

Alberto Salvo†
Department of Economics
National University of Singapore

September 2014

1 Dynamic type evolution

We provide examples, from the data, of the dynamic updating process. We consider two transitions, both for São Paulo Metro (region 4). The first transition, from $t = 1$ to $t = 2$, features upward mobility and, unusually in the data (yet we need to allow for this), a slight flow of residents out of the region (“net urban-to-rural migration”). The second transition, from $t = 10$ to $t = 11$, features upward mobility and the rural-to-urban migration that is prevalent in the data. We illustrate these transitions at the estimated model parameters θ^*. We also comment on the robustness of our estimates to the baseline mobility Assumptions 1 and 2.

Region 4, $t = 1$ to $t = 2$. The initial type-distribution vector is

$$\begin{aligned}\mathcal{F}_{g=4,t=1} &= \left\{F_{EA^{\mathcal{A}},4,1}, F_{EA^{\mathcal{B}},4,1}, F_{EA^{\mathcal{O}},4,1}, F_{NA^{\mathcal{A}},4,1}, F_{NA^{\mathcal{B}},4,1}, F_{NA^{\mathcal{O}},4,1}, F_{P^{\mathcal{A}},4,1}, F_{P^{\mathcal{B}},4,1}, F_{P^{\mathcal{O}},4,1}\right\} \\ &= \{0.255, 0.029, 0.361, 0, 0, 0, 0.048, 0.009, 0.299\}\end{aligned}$$

As explained, the last element, for instance, is the product of (region 4’s) poor household count in $t = 1$ (observed from combining IBOPE and IBGE) and the share of the outside option among region 4’s **DE** households (calculated from the HEX 95/96), divided by the total household count (IBOPE/IBGE), i.e., $1346585 \times 0.84093/3789771 \simeq 0.299$. From $s_{j,r,g=4,t=1}(\theta^*)$ (see (4)), we obtain the mass of households for each of the nine types who choose to consume premium, generic, or no soda. For example, the premium soda share among established affluent households who in the previous period consumed premium soda is $\sum_{j\in\mathcal{A}} s_{j,EA^{\mathcal{A}},g=4,t=1}(\theta^*) \simeq 94\%$. In contrast, the premium soda shares among established affluents who previously consumed generic soda and no soda are, respectively, 3% and 20%. Thus, since the established affluent population is constant over time (at 2443186), the number of established affluent households going into $t = 2$ in the premium-soda prior-consumption state is (in thousands, hereafter) $3790\,(0.255 \times 0.94 + 0.029 \times 0.03 + 0.361 \times 0.20) \simeq 1192$.

*aloneiz@mscc.huji.ac.il
†albertosalvo@nus.edu.sg

As for mobility, according to IBOPE/IBGE, the socioeconomic distribution of households evolves from (**ABC**,**DE**) = $(2443, 1347)$ in $t = 1$ to $(2511, 1269)$ in $t = 2$. It follows that, in $t = 2$: (i) $2511 - 2443 = 68$ households are newly affluent; (ii) $(2443 + 1347) - (2511 + 1269) = 10$ households migrated out of the urban area (again, this rarely happens in the data); and (iii) 1269 households are poor. Following Assumption 1, the 68 upwardly mobile households entering $t = 2$ are assigned to states in proportion to the choices of poor households in $t = 1$ among premium, generic and no soda (where these proportions are calculated as illustrated for established affluents, for which a proportion $1192/2443 \simeq 49\%$ chose premium rather than generic or no soda). These counts (summing 68) are deducted from the $t = 1$ poor population (1347) that is transitioning to $t = 2$ in proportion to the poor's choices across brand types. Similarly, following Assumption 2 (Migration), the 10 households leaving the city are dropped from the counts of the poor (totaling $1347 - 68$) in proportion to the poor's choices across brand types.

Region 4, $t = 10$ to $t = 11$. This example highlights mobility. The type-distribution vector following choices made in $t = 9$ and mobility into $t = 10$ is

$$\mathcal{F}_{g=4,t=10} = \{0.250, 0.109, 0.279, 0.003, 0.007, 0.137, 0.000, 0.002, 0.214\}$$

Having updated from $t = 1$, the history of choices and mobility now determines the distribution of states across each socioeconomic group. At $t = 10$, the generic-to-premium ratio is $0.109 : 0.250 = 0.4$ among the established affluent (see Table 5). From IBOPE/IBGE data, the mass of households by socioeconomic group (in thousands) in $t = 10$ is computed as: 2443 established affluent (this stays constant), 560 newly affluent and 826 poor (see Figure 4; $t = 10$ is the Jun-98/Jul-98 bimonth). The evolution of (**ABC**,**DE**) from $(3003, 826)$ in $t = 10$ to $(3060, 784)$ in $t = 11$ implies that: (i) the newly affluent count grows by 57 (to 617); (ii) 16 migrants arrive at the city and join the ranks of the poor; and (iii) the poor count drops by $57 - 16 = 42$ (to 784). The 57 upwardly mobile households making choices with newly affluent status in $t = 11$ belong to states in proportion to the $t = 10$ choices of the poor they left behind (Assumption 1). The 16 migrants who are new to the city are assigned to the no-soda state (Assumption 2).

Robustness to Assumptions 1 and 2. Our results are robust to alternative assumptions, namely: (i) modifying Assumption 1 to assign households moving up from poor to newly affluent status to prior-consumption states in proportion to the previous-period soda choices of the newly affluents they are joining, rather than the poor they are leaving behind[1] (and analogously with respect to households moving down from newly affluent to poor status, based on previous-period choices by the poor); and (ii) modifying Assumption 2 to assign households moving to urban areas to prior-consumption states in proportion to the previous-period choices of the city-dwelling poor they are joining. For example, under (ii), $(\alpha_{EA}, \alpha_{NA}, \alpha)$ and λ are estimated, respectively, at $(3.62, 1.87, -5.76)$ and 4.21 (with standard errors of $(1.42, 1.45, 1.47)$ and 0.26), very close to baseline estimates (Table 4). Estimates under variant (i) are also very close to baseline.

[1] The exception is the first transition, from $t = 1$ to $t = 2$, in which the newly affluent are a random sample of the poor as, by definition, there are no newly affluents in $t = 1$.

2 The estimation algorithm

We classify the parameters θ into "linear" and "non-linear" parameters, $\theta_1 = \left\{\beta, \alpha\right\}$ and $\theta_2 = \left\{\alpha_{EA}, \alpha_{NA}, \lambda\right\}$, consistent with familiar terminology from the literature. Given any generic value for the non-linear parameters θ_2, steps 1 to 5 of the algorithm below yield an $N \times 1$ vector $\delta\left(\theta_2\right)$, containing the base utilities for every brand in every region-period market ($N = 9 \cdot 7 \cdot 57$). As noted in Section III, conditioning on the full parameter vector $\theta = (\theta_1, \theta_2)$, one obtains an $N \times 1$ vector of demand unobservables by subtracting the systematic portion of the base utility from δ_{jgt}, i.e., $\xi_{jgt} = \delta_{jgt} - x'_{jgt}\beta - \alpha p_{jgt}$. Stacking these unobservables, we obtain:

$$\xi(\theta) = \delta\left(\theta_2\right) - X\theta_1$$

where the $N \times K_1$ matrix X contains the K_1 base-utility covariates (including price), and let K_2 denote the dimension of θ_2. Now let Z denote a $N \times L$ matrix of instruments containing all covariates in X but price, as well as excluded instruments, where $L > K_1 + K_2$. Writing $W = (Z'Z)^{-1}$, the GMM objective is defined by:

$$Q_N(\theta) = \xi(\theta)' ZWZ' \xi(\theta)$$

Computation time can be reduced substantially by noting (see BLP 1995, Nevo 2000) that, conditional on θ_2, there is a closed-form solution for θ_1 that minimizes the objective:

$$\theta_1^*\left(\theta_2\right) = \left(X'ZWZ'X\right)^{-1} X'ZWZ'\delta\left(\theta_2\right)$$

This allows us to maximize the objective by searching only over values for θ_2. At every guess $\tilde{\theta}_2$ for the non-linear parameters, the GMM objective is evaluated via the following steps:

1. For every region $g = 1, ..., 7$, and period $t = 1$, given $\mathcal{F}_{g1}$ and $\tilde{\theta}_2$, use the BLP contraction mapping to solve for the unique vector of base utilities that matches observed aggregate market shares with those predicted by the model.

2. For every region $g = 1, ..., 7$ and household type $r = 1, ..., 9$, use equation (4), the base utilities recovered in step 1, and $\tilde{\theta}_2$, to predict the shares of type-r households who consume premium brands, generic brands or no soda in period $t = 1$.

3. For every region $g = 1, ..., 7$, use the shares obtained in step 2, data on aggregate social mobility and migration, and Assumptions 1 and 2, to obtain next period's type-distribution vector, $\mathcal{F}_{g2}$ (recall Section III.B).

4. Repeat steps 1-3 for periods $t = 2, ..., 57$.

5. Stack the base utilities for all brands in all regions and periods in the $N \times 1$ vector $\delta\left(\tilde{\theta}_2\right)$, and evaluate the GMM objective at $\tilde{\theta}_2$, as explained above.

3 Data

Figure A1 reports variation over the sample period in: (i) the intensity of premium brands' media advertising (left panel, summed cross brands, meaned across regions); and (ii) retail distribution for the main premium brand, Coke, and for generic brands (right panel, meaned across regions). The sources are McCann-Erickson and Nielsen, respectively. The fluctuation in premium soda advertising is considerably larger than that of premium soda quantities sold and does not share as seasonal a pattern (Figure 1). The proportion of self-service outlets that were stocking at least one generic brand on the day of Nielsen's audit already stands at a fairly high 83% in the first period and grows to 98% by mid 1999, whereas the Coke brand is stocked by close to 100% of stores all along. Our baseline specification controls for such brand-level variation in "visibility" both in the media and at the store.

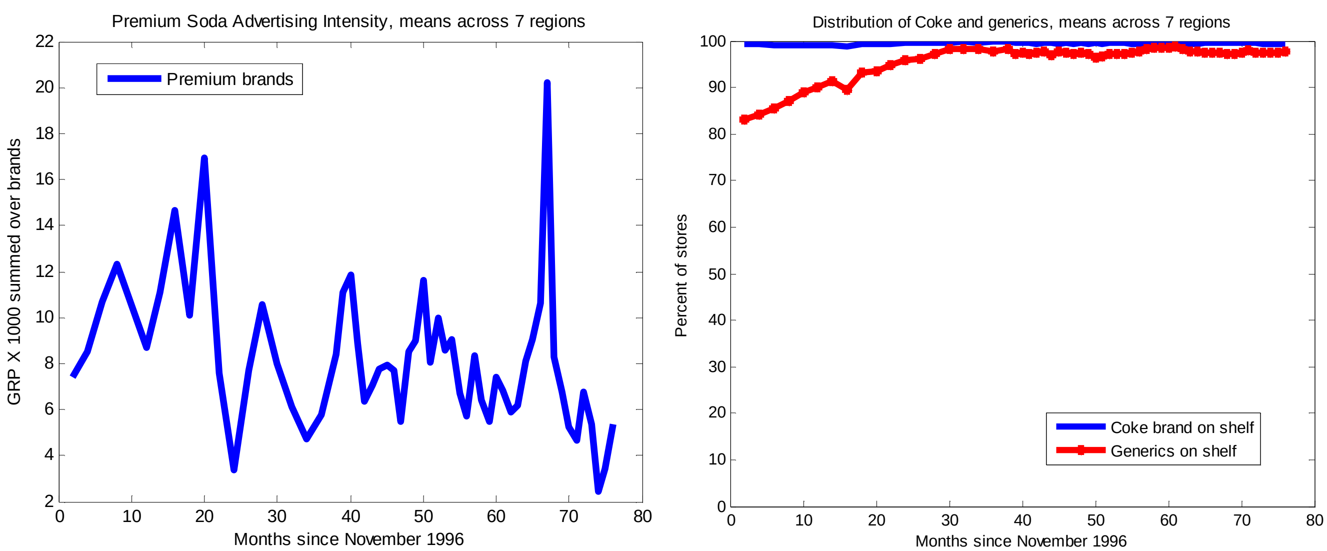

Figure A1: The evolution of premium soda media advertising intensity (in monthly Gross Rating Points, GRP, ×1000) and the proportions of stores with specific brands in stock (in percent), in the period Dec-96 to Mar-03. GRP and percentages shown are means across the seven Nielsen regions. Sources: McCann-Erickson, Nielsen.

The temperatures we include as additional demand shifters are region-specific monthly averages, sourced from the National Institute of Meteorology. Input prices that we include among demand instruments are the Fundação Getúlio Vargas's wholesale price for refined sugar (the 'IPA-OG açúcar') and price index for transportation fuel (the 'IPA-OG combustíveis e lubrificantes'), besides regional high-voltage electricity prices ('classe industrial') obtained from the National Agency for Electrical Energy. We inflation-adjust all nominal prices—for soda and for inputs—using a consumer price index (the 'IPC-br') published by the Fundação Getúlio Vargas. For perspective, the CPI has averaged +7.8% per year over the sample period.

The remaining part of this subsection explains how we combine IBOPE's LatinPanel survey with IBGE's annual household survey (the PNAD, 'Pesquisa Nacional por Amostra de Domicílios') to produce household counts by socioeconomic standing. We also detail how we obtain the first period's type-distribution vector, $\mathcal{F}_{g1}$, from IBGE's 1995/96 urban household expenditure survey (the 'Pesquisa de Orçamentos Familiares').

Data on aggregate social mobility. From IBOPE's LatinPanel we observe, for every region-year pair, what proportion of urban households belong to each of two socioeconomic groups, **ABC** or **DE**. IBOPE's regions map directly into Nielsen's seven regions, with the exception of region 1 (northeastern states less Maranhão and Piauí) for which IBOPE's coverage includes all northeastern and northern states. Since Maranhão, Piauí and the North comprise the country's least urbanized and least populated area, we simply take IBOPE's urban distributions for the Northeast/North as representative of urban households in Nielsen's region 1. IBOPE's survey through 2002 was representative of all municipalities with populations of at least 20,000, and their coverage was expanded in 2003 to represent municipalities with populations exceeding 10,000.

We obtain household counts from IBGE's annual household surveys (PNAD). These cover households, both urban and rural, in all 27 states of the country. For perspective, 115,654 households were sampled in 1999. We use the weights provided to expand the representative sample to the universe of households. We consider only households residing in urban areas and in states within each Nielsen region. For example, for region 1, we sum the number of urban households across all states in the Northeast less Maranhão and Piauí.

We then multiply, for each Nielsen region and year, the IBOPE socioeconomic proportions of urban households by the IBGE urban household counts. To increase the frequency of the resulting panel from annual to monthly periods (or bimonthly periods, thus matching the frequency of Nielsen's point-of-sale audits), we linearly interpolate from September of one year to September of the following year, noting that September is the IBGE PNAD's annual "month of reference."

Data on household-level brand choices. IBGE's HEX 95/96 surveyed 16,013 households in 11 large metropolitan areas across the country. Carvalho Filho and Chamon (2012) discuss this survey in detail. Over a reference period of one week falling between October 1995 and September 2006, the soft drink expenditure in R$ for consumption inside the home was recorded for each household, detailed by soda brand(s) purchased. We then classified the following brand descriptions and codes as "premium" brands: *Coca-Cola* (9301); *Pepsi* (9302); *Guaraná* (9303); *Fanta laranja, uva, limão* (9304); *Soda limonada* (9307); *Mirinda* (9308); *Sukita* (9315); *Pop laranja* (9316); and *Refrigerante água tônica* (9349). Examples of coded brand descriptions that we classified as "generic" brands are: *Refrigerante tubaína* (9318); *Refrigerante laranja exceto Fanta, Sukita, Pop, Crush* (9339); *Refrigerante cola exceto Coca-Cola e Pepsi-Cola* (9340); *Refrigerante cajú qualquer marca* (9346); and *Refrigerante Goianinha* (9355). Of the 16,013 households, 10,172 (or 64% of households) were recorded as making no soda purchases, 4,465 households (28%) purchased only brands that we can confidently identify as premium, 310 households (2%) purchased only brands that we identify as generic, and 236 households (only 1%) simultaneously purchased brands that we identify as premium and brands that we identify as generic. This observation justifies our modeling of soda-consuming households at each point in time as either premium or generic shoppers, but not "hybrids."

We deemed four soda descriptions to be ambiguous with regard to brand type: *Refrigerante água natural* (9310); *Refrigerante gasosa* (9319); *Refrigerantes não especificado* (9335); and *Refrigerante dietético* (9360). We need to assign the soda expenditure of the remaining 830 soda-purchasing households (5% of the survey sample) to either premium brands or generic brands. These are households whose soda expenditure we cannot entirely identify by brand type, such as a household purchasing R$ 4 of Coke (9301) and R$ 2 "Soda not specified" (9335). To do this, we first designate as premium the "brand-unidentifiable" soda expenditure portion (R$ 2 in the example) for those households whose identifiable-premium expenditure share of soda exceeds 50% (Coke's 67% share in the example) *and* identifiable-generic expenditure share of soda is less than 10% (0% in the example).

Similarly, we assign to generic the unidentifiable soda expenditure portion for those identified-generic-dominant households. Finally, the soda expenditure portion that for a remaining 614 households is still not assigned to a brand type—e.g., a yet-to-be-assigned R$ 4 purchase of "Soda not specified" (9335)—is allocated among premium and generic expenditure: (i) in proportion to the (identified or designated) premium versus generic expenditure shares within the household; or (ii) if none of the household's soda expenditure is identifiable (say the household in this second example purchased only "Soda not specified"), the allocation is done in proportion to the premium versus generic expenditure shares across households in the same socioeconomic group and metropolitan area.

We use balance sheet data to classify households according to socioeconomic standing, **ABC** or **DE**, as described in Section I. We use the weights provided to expand the representative sample to a universe of 12.5 million households across the 11 metropolitan areas. To calculate household-level premium and generic quantities, we divide HEX 95/96 expenditures on premium and generic soda by Nielsen's region-specific share-weighted mean prices for premium and generic brands, respectively, in period $t = 1$ (Dec-96/Jan-97). We then aggregate premium (resp., generic) quantities across the universe of households in each socioeconomic segment and in the surveyed metropolitan areas for each Nielsen region g (e.g., the cities of Recife, Fortaleza and Salvador in the Northeast, $g = 1$).

The premium (resp., generic) soda shares among the initial masses of established affluent and poor households are calculated analogously to how we define $s_{jgt} = q_{jgt}/\mathcal{M}_{gt}$ in Section III, i.e., taking market size (in our base specification) as six liters per household per week times the number of weeks in period $t = 1$. Combining these premium versus generic (versus no soda) shares by socioeconomic group with first-period household counts by socioeconomic group (as per above), yields $F_{EA^{\mathcal{A}},g1}, F_{P^{\mathcal{A}},g1}$ and $F_{EA^{\mathcal{B}},g1}, F_{P^{\mathcal{B}},g1}$ (recall that $F_{NA^{\mathcal{A}},g1} = F_{NA^{\mathcal{B}},g1} = 0$). We consider only soda purchases that were recorded as being for the household's inside-the-home consumption (rather than recorded as "individual consumption") and at stores coded as *Supermercado* (1), *Hipermercado* (2), *Padaria* (3), *Lanchonete* (11), and *Mercado & Central de Abastecimento* (26), in view of the mapping to Nielsen's self-service channel (stores with checkouts).

Two points are noteworthy. First, the HEX survey suggests that urban household size does not vary significantly across socioeconomic group. Mean household sizes are 3.64 for **ABC** (with a standard deviation of 1.58) and 3.76 for **DE** (s.d. 1.99). Second, while the HEX shares that enter the initial conditions $\mathcal{F}_{g1}$ are calculated following the market share definition of Section III, one should note that the shares reported in Table 2 are extensive margins of soda consumption, i.e., the proportions of households who purchase any soda quantity. As for intensive margins, the modal intensity of soda consumption, conditional on positive consumption, is 2 liters per household per week regardless of the socioeconomic group and the region. One can intuitively interpret this pervasive intensive margin as "one 2-liter family-size bottle of soda that is brought to the table every week."

We performed all manner of "consistency checks," where applicable, to ensure that the data were consistent across the different sources. For example, among households residing in the three metropolitan areas in the Northeast that were surveyed in the HEX, premium and generic market shares amount to 12.5% and 0.4%, respectively.[2] These HEX shares of 12.5% and 0.4% are similar to the Nielsen market shares of 12.3% across premium brands and 0.3% for generics in the Dec-96/Jan-97 bimonth (soda sold in family-size bottles through self-service outlets in the Northeast). By way of another example, the HEX 95/96 projects the universe of households for region 3 (the Rio de Janeiro metropolitan area) at 2.96 million (Table 2), to be compared to 2.64 million households projected for Dec-96 in the IBGE PNAD (noting that Nielsen's region

[2] These shares are as defined in Section III. Shares grow to 26.6% and 0.7%, respectively, if we condition on **ABC** households (these are similar to the extensive margins reported in Table 2).

3, which we adopt for the IBGE household counts, excludes some peripheral villages around the city of Rio de Janeiro). Further, using the HEX 95/96's balance sheet data, we assigned **ABC** socioeconomic status to 55% of region 3's households (Table 2), whereas the IBOPE data suggest that at that time 57% of region 3's households were **ABC**.

4 Robustness

We briefly describe some of the alternative specifications (on top of the alternative mobility assumptions discussed in section 1 of this online Appendix) that we have estimated to confirm the validity and robustness of our baseline results. Estimates of these model variants are available upon request.

Market size. Our baseline model defines market potential as six liters per week, interpreted as 3 meals/week in which a 2-liter family-size bottle of soda might be brought to the table. Estimated price sensitivities and the persistence parameter hardly vary as we vary the number of meals per week between 2.8, 2.7, ... , 3.7. Beyond this range, all the way from 2.0 to 4.0 meals/week, estimated $(\alpha_{EA}, \alpha_{NA}, \alpha)$ vary more, but our estimate for λ is very stable, between 4 and 5.

Dynamic dependence. Specifications that we implemented, each addressing alternative mechanisms than the one we wish to highlight, include: (i) allowing "loyalties" to form for the flagship premium brands Coke (including Diet Coke), Guaraná Antarctica, Fanta, or Pepsi; for example, consuming Pepsi in this period increases one's utility from consuming Pepsi in the next period—but not another brand—by λ; and (ii) allowing loyalty to form only for the Coke (including Diet Coke) brand. To illustrate, model (i) has five loyalty states which, interacted with 3 socioeconomic groups, implies 15 household types (and we must modify the initial conditions from the HEX accordingly). Brand loyalty is estimated to be strong and significant under alternative models (i) and (ii), leading to aggregate own-price elasticities that appear too low in magnitude, namely, -0.6 and -1.1 for Coke under (i) and (ii), respectively (compared to -1.7 in Table 6). In general, estimated persistence parameter(s) under these alternative models are large and significant, but do not provide as strong a justification for Coca-Cola's mid 1999 price cut.

Other specifications. A more general model allowed the prior-period consumption effects to vary in magnitude for premium ($\lambda^{\mathcal{A}}$) and generic ($\lambda^{\mathcal{B}}$). Estimated persistence parameters are $\lambda^{\mathcal{A}} = 4.67$ (s.e. 0.33) and $\lambda^{\mathcal{B}} = 2.90$ (s.e. 0.49), and price sensitivity is similar to baseline, namely $(\hat{\alpha}_{EA}, \hat{\alpha}_{NA}, \hat{\alpha}) = (3.71, 2.12, -5.94)$ with s.e. $(1.71, 1.61, 1.75)$. Further, our estimates are robust to: (i) dropping media advertising ($\hat{\lambda} = 4.31$ with s.e. 0.25, $(\hat{\alpha}_{EA}, \hat{\alpha}_{NA}, \hat{\alpha}) = (4.74, 2.94, -6.97)$ with s.e. that almost double compared to Table 4); and (ii) dropping retail distribution ($\hat{\lambda} = 4.93$

with s.e. 0.35, $(\hat{\alpha}_{EA}, \hat{\alpha}_{NA}, \hat{\alpha}) = (3.07, 2.67, -5.33)$ with s.e. that are similar to Table 4). We also tested robustness with regard to: (i) initial HEX 95/96 shares (namely, expanding the HEX outlet codes that map to Nielsen's stores with checkouts); and (ii) defining market share by the extensive margin once the region-specific intensive margin, as observed in the single cross-section of household-level data (HEX 95/96), is fixed over time.

5 Variable profit

Our back-of-the-envelope calculation of variable profit considers the three-year period between April 2000 and March 2003. We assume that the premium sellers' net sales price is 35% of the price Nielsen observes on the shelf, which is paid by the end consumer. (See Ambev 2003 and Salvo 2009 for a discussion of the very high taxes incurred along the formal vertical chain, as well as vertical relations. We also base our calculations on interviews with an executive at a premium seller.) Thus, the observed shelf price of R$ 0.913 / liter (sales weighted across premium brands, averaged over the three years) corresponds to a net sales price for Coca-Cola/Ambev of R$ 0.320 / liter, net of sales tax, retail margin, and distribution costs. Had the premium sellers not cut prices in mid 1999, we assume that this price would have been proportionately higher, at R$ 0.385 / liter. Based on Ambev (2003), we take the "cost of goods sold" as R$ 0.199 / liter. We note that the real prices of sugar, plastic, electricity, labor, and fuel were quite stable between 2000 and 2002 (in general, they began rising at the end of our sample period, in 2003). The variable profit margins for the Coca-Cola/Ambev "systems" are thus R$ 0.120 / liter with the observed price cut and R$ 0.186 / liter under the counterfactual of no price cut. Multiplying by the premium sellers' observed and counterfactual quantities sold over this three-year period (namely, 7.2 billion liters observed; 4.3 bi liters counterfactual "no price cut" under the Brand Type Persistence model) yields the variable profits stated in the text (respectively, R$ 861 million, R$ 807 million).

6 Monte Carlo experiments

Data generating process. Demand for soda follows the state-dependent household-level choice model developed in Section III, namely indirect utility (2) with baseline persistence feature (3). We take as true parameters the point estimates reported in Table 4—these are reproduced in column I of Table A1. We design each simulated dataset to have the same dimensions as our empirical dataset: 9 brands (8 of which are premium), 7 regions and 57 time periods. This enables us to take covariates x_{jgt}, the first-period distribution of types $\mathcal{F}_{g1}$, and the evolution of each region's household population by socioeconomic group as observed in the data (see definitions in Section III). Assumptions 1 and 2 ("orthogonality") dictate how demographic shifts interact with previous-period consumption. Prices are simulated according to:

$$p_{jgt} = \begin{cases} \Lambda_t^{prem} c_{jt}^{prem} + \rho \xi_{jgt} \text{ if } j \in \mathcal{A} \\ c_{jt}^{gen} \text{ if } j \in \mathcal{B} \end{cases}$$

where c_{jt}^{prem} and c_{jt}^{gen} are marginal costs for premium and generic brands, respectively, that are flat in output and vary across brands and over time, but not across regions (one can relax this); Λ_t^{prem} is a time-varying price markup over marginal cost for premium brands (this can also be made to vary across brands and regions); unobserved taste shocks ξ_{jgt} are i.i.d. across brands (including generics), regions and time; and $0 \leq \rho < 1$ is a pass-through ratio from utility shocks for premium brands to prices. Define marginal costs as the time-varying price of inputs W (e.g., sugar) times a cost efficiency (inverse productivity) parameter τ that varies by brand type (premium or generic) and over time, plus a brand-and-time varying disturbance term u:

$$\begin{aligned} c_{jt}^{prem} &= W_t \tau_t^{prem} + u_{jt} \\ c_{jt}^{gen} &= W_t \tau_t^{gen} + u_{jt} \end{aligned}$$

Specifically, the simulations reported in Table A1—except column V—consider price variation that is inspired by (is "comparable" to) what we observe in the real data: (i) premium brands' markup $\Lambda_t^{prem} \sim N(2, 0.02^2)$ until April-May 1999 and, following the 20% price cut, $\Lambda_t^{prem} \sim N(1.6, 0.02^2)$ thereafter; (ii) factor price $W_t \sim U(0.85, 1.15)$; (iii) premium brands' cost efficiency τ_t^{prem} equal to 0.55 throughout the sample period; (iv) generics' cost efficiency τ_t^{gen} equal to 0.9 in the first period, decreasing linearly to 0.6 in August 2000, and constant thereafter; and (v) brand cost disturbance $u_{jt} \sim N(0, 0.01^2)$.

By contrast, the simulation reported in column V, marked "less simulated data variation versus real": (i) drops the premium price cut, simulating the premium brands' markup according to $\Lambda_t^{prem} \sim N(2, 0.02^2)$ throughout; (ii) drops the fringe's price decline, setting the generics' cost efficiency τ_t^{gen} already at 0.6 from the first period; and (iii) freezes each region's household population by socioeconomic group (EA, NA, P) from the second period on.

Completing the description of the simulated datasets, we consider a pass-through ratio $\rho = 0.3$ and model utility shocks $\xi_{jgt} \sim N(0, \sigma^2)$ with varying orders of magnitude of variance, as reported in each column of Table A1. Notice that our experiments do not rely on the established firms pricing optimally, and that premium prices are endogenous since the established firms pass through a proportion of brand-market specific taste shocks to prices. Because taste shocks are i.i.d. and prices correlate across regions through a common cost (and markup) structure, prices in one region are a valid instrument for prices in another region.

Estimation. To estimate using simulated data, we follow the estimation procedure proposed in Section III. This includes adopting the same three instrument classes, namely: (i) the price of inputs, (ii) the contemporaneous mean price for a brand in the other regions, and (iii) a dummy variable indicating periods after July 1999 interacted with brand-region fixed effects. Since we calculate household-type specific brand shares analytically, there is no sampling variation in the logit shock ϵ_{ijgt}. For every simulation $\omega = 1, ..., \Omega$, we complete the experimental—i.e., aggregate—dataset and make inference using the true (baseline) demand model (see columns II to V), as well as the No Persistence model variant (column VI).

Results. Table A1 velow reports results. With $\sigma = 0.001$, the estimated baseline model recovers the true parameters to at least 2 decimal places (column II). Precision falls but is still quite high, particularly for $\hat{\lambda}$, when $\sigma = 0.1$ (column IV). Estimates using simulated data that exhibits less rich variation than what we observe empirically—namely, where we shut down the premium price cut, the generic price decline and the socioeconomic transitions—are substantially more noisy (column V compared to column IV). Estimates using the No Persistence model variant (column VI compared to column IV) are noisier and, importantly, price sensitivity appears biased downward (for brevity, the bottom of the table reports aggregate own-price elasticities only).

Table A1: Monte Carlo Experiments

	I	II		III		IV		V		VI	
Model		Baseline		Baseline		Baseline		Baseline		NP	
Simulated vs. real		'Comparable'		'Comparable'		'Comparable'		'Less'		'Comparable'	
σ		0.001		0.01		0.1		0.1		0.1	
	True	Median	(SD)	Median	(SD)	Median	(SD)	Median	(SD)	Median	(SD)
Price sensitivity											
α_{EA}	**3.56**	3.56	(0.00)	3.57	(0.04)	3.55	(0.38)	3.20	(0.71)	4.71	(0.90)
α_{NA}	**1.80**	1.80	(0.00)	1.80	(0.04)	1.80	(0.37)	0.67	(23.80)	3.15	(0.62)
α	**-5.70**	-5.70	(0.00)	-5.71	(0.04)	-5.66	(0.42)	-5.24	(0.65)	-5.55	(0.95)
Combinations:											
$\alpha + \alpha_{EA}$	**-2.14**	-2.14	(0.00)	-2.14	(0.01)	-2.09	(0.06)	-2.04	(0.14)	-0.76	(0.15)
$\alpha + \alpha_{NA}$	**-3.90**	-3.90	(0.00)	-3.90	(0.01)	-3.85	(0.14)	-4.34	(23.82)	-2.31	(0.59)
$\alpha_{EA} - \alpha_{NA}$	**1.76**	1.76	(0.00)	1.76	(0.01)	1.76	(0.10)	2.27	(23.83)	1.47	(0.56)
Persistence											
λ	**4.20**	4.20	(0.00)	4.20	(0.01)	4.16	(0.09)	4.18	(0.15)	–	–
Other Effects											
Constant	**-3.84**	-3.84	(0.00)	-3.84	(0.01)	-3.82	(0.09)	-3.87	(0.17)	-2.47	(0.20)
Temperature	**2.91**	2.91	(0.00)	2.91	(0.01)	2.85	(0.07)	2.82	(0.07)	1.66	(0.21)
Advertising×Region 1	**0.44**	0.44	(0.00)	0.44	(0.01)	0.43	(0.07)	0.45	(0.07)	0.22	(0.08)
Advertising×Region 2	**0.25**	0.25	(0.00)	0.25	(0.01)	0.25	(0.08)	0.25	(0.08)	0.21	(0.08)
Advertising×Region 3	**0.00**	0.00	(0.00)	0.00	(0.01)	0.01	(0.08)	0.02	(0.08)	0.02	(0.09)
Advertising×Region 4	**0.40**	0.40	(0.00)	0.40	(0.01)	0.40	(0.09)	0.39	(0.09)	0.48	(0.08)
Advertising×Region 5	**0.32**	0.32	(0.00)	0.32	(0.01)	0.33	(0.08)	0.32	(0.08)	0.36	(0.09)
Advertising×Region 6	**0.66**	0.66	(0.00)	0.66	(0.01)	0.67	(0.09)	0.67	(0.09)	0.82	(0.09)
Advertising×Region 7	**0.49**	0.49	(0.00)	0.49	(0.01)	0.47	(0.07)	0.48	(0.07)	0.48	(0.09)
Distribution×Region 1	**3.10**	3.10	(0.00)	3.10	(0.01)	3.11	(0.06)	3.12	(0.06)	4.12	(0.12)
Distribution×Region 2	**3.70**	3.70	(0.00)	3.70	(0.01)	3.69	(0.10)	3.69	(0.10)	3.62	(0.11)
Distribution×Region 3	**4.09**	4.09	(0.00)	4.09	(0.01)	4.08	(0.14)	4.09	(0.13)	4.59	(0.21)
Distribution×Region 4	**2.72**	2.72	(0.00)	2.72	(0.01)	2.73	(0.12)	2.72	(0.12)	2.72	(0.15)
Distribution×Region 5	**3.33**	3.33	(0.00)	3.33	(0.01)	3.34	(0.12)	3.33	(0.12)	3.31	(0.11)
Distribution×Region 6	**1.24**	1.24	(0.00)	1.24	(0.01)	1.25	(0.07)	1.24	(0.07)	1.41	(0.08)
Distribution×Region 7	**1.11**	1.11	(0.00)	1.11	(0.01)	1.11	(0.06)	1.11	(0.06)	0.97	(0.12)
Time Trend Region 1	**-0.37**	-0.37	(0.00)	-0.37	(0.00)	-0.36	(0.03)	-0.38	(0.02)	0.03	(0.15)
Time Trend Region 2	**0.13**	0.13	(0.00)	0.13	(0.00)	0.13	(0.03)	0.13	(0.02)	0.24	(0.21)
Time Trend Region 3	**-0.33**	-0.33	(0.00)	-0.33	(0.00)	-0.32	(0.02)	-0.32	(0.02)	0.29	(0.13)
Time Trend Region 4	**-0.66**	-0.66	(0.00)	-0.66	(0.00)	-0.65	(0.03)	-0.65	(0.02)	-0.20	(0.10)
Time Trend Region 5	**0.02**	0.02	(0.00)	0.02	(0.00)	0.03	(0.04)	0.02	(0.02)	0.52	(0.19)
Time Trend Region 6	**-0.05**	-0.05	(0.00)	-0.05	(0.00)	-0.05	(0.03)	-0.05	(0.02)	0.37	(0.16)
Time Trend Region 7	**-0.07**	-0.07	(0.00)	-0.07	(0.01)	-0.07	(0.06)	-0.08	(0.03)	0.47	(0.22)
Aggregate Own-Price Elasticities											
Coke		-1.33	(0.02)	-1.33	(0.02)	-1.31	(0.03)	-1.43	(0.04)	-0.83	(0.11)
Guaraná Antarctica		-1.98	(0.02)	-1.98	(0.02)	-1.94	(0.05)	-2.09	(0.07)	-0.96	(0.14)
Fanta		-2.02	(0.02)	-2.02	(0.02)	-1.99	(0.05)	-2.15	(0.07)	-0.97	(0.14)
Pepsi		-2.06	(0.02)	-2.06	(0.02)	-2.02	(0.05)	-2.19	(0.07)	-0.98	(0.14)
Generic		-0.75	(0.01)	-0.75	(0.01)	-0.75	(0.01)	-0.55	(0.02)	-0.77	(0.09)

Medians and standard deviations (in parentheses) of estimated parameters are taken over $\Omega = 50$ simulations. Reported elasticities are means across region-and-time markets.