Supplement to Push-Me Pull-You: Comparative Advertising in the OTC Analgesics Industry. **For Online Publication**

April 2016

Appendix A: Dataset Construction in Detail

A.1 Sales Dataset

We take the original dataset and we follow these steps:

1. We drop all the products that are analgesics but do not come in the form of pills or are unconventional analgesics (e.g., pain relief patches).

2. We drop Pamprin and Midol, which have less than 1% of the inside market - much less than other brands; in addition, they specialize only in menstrual pain. We also drop all of the medication that is classified as cold medicine, since it is highly seasonal.

3. We determine the active ingredient for each product: acetaminophen, ibuprofen, naproxen, or aspirin.

4. We assign a number of milligrams to each product, according to the strength of its primary active ingredient. To do so, we combine the descriptive data in the Nielsen dataset with the data of milligrams of a specific active ingredient in a specific formula. In the case of Ibuprofen- and Naproxen Sodium- based pain relievers, the assignment was straightforward, since these OTC products can come only in 200mg (for Ibuprofen) and 220mg (for Naproxen Sodium). In the case of Aspirin and Acetaminophen, the situation is more delicate, since these products can come in varying strengths and as a combination with other analgesic agents. Therefore, we consider whether the product is of regular strength, extra strength, body and back pain (which includes caffeine), and rapid headache; whether the product is for rapid release; whether the product is for children; whether the product has a sleepaid; whether it is for arthritis; for migraine, for menstrual purposes; or for sinus headache.

5. For a certain analgesic drug to be sold as an OTC drug, the FDA requires that the daily (24 hours) dosage does not exceed a certain threshold (the thresholds are different for different active ingredients. For example, for acetaminophen the daily dosage is 4000 mg. of this active ingredient). Thus, we create a variable that indicates the maximum amount of pills allowed in 24 hours by the FDA regulation.

6. We deflate the prices of the pills by the CPI (January 2000=100).

7. We define the *market size, M,* for OTC analgesic products as the US population 18 years or older minus the estimated number of people who buy pain medication at Wal-Mart, a store that does not provide information on the sales of products.

8. We construct a measure of a *serving* of pain medication, or a *pain episode*, so that we can aggregate across different package sizes and across different medication strengths. We express each product's sales as the number of people whose pain could be relieved by that product for a period of three days, which is the average number of pain days per month in the population.[1]. To this end we assigned to each analgesic product in the sales dataset the strength of its active ingredient in milligrams and derived the maximum number of pills that a consumer can take for OTC analgesics consumption over 72 hours as defined by the FDA and required to be listed on the labels (e.g. 9 in the case of Aleve, and from 18 to 36 for Tylenol, depending on the ACT formulation). This we refer to as an episode of pain.

9. We take the number of pills in a pack and multiply by the number of packs sold. We divide this number by 3 (the average number of sick days per month) and we divide the result by the maximum number of pills allowed by the FDA to obtain the number of servings sold for each type of pill in a month. This is how we compute the market share for each product.

10. As discussed in the text, we do the same exercise for the generic products, which are differentiated only by their active ingredient. Thus, we assume the generic products are provided by a competitive fringe.

11. Some of the firms in our analysis are multi-brand firms. Motrin and Tylenol are owned by McNeil, and Aleve and Bayer are owned by Bayer. We treat each brand as making independent decisions.[2]

[1] This information is from the Morbidity and Mortality Weekly Report, Centers for Disease Control and Prevention, Feb 27, 1998/47(07); 134-140.

[2] This is not a problem at all for the self-promotion equation, which is exactly the same if we allow firms to behave as multi-brand firms that maximize joint pay-offs. However, the comparative ad equation would be modified to include cross-brand effects. This would require the estimation of a large number of additional diversion ratios with the same number of observations, which in exploratory work resulted in many diversion ratios being imprecisely estimated. We therefore treat brands as independent divisions maximizing brand profits, modulo the imposition that they do not attack sibling products as concurs with the data in this respect.

A.2 Advertising Dataset

1. When coding the ads, a few things need to be kept in mind: the same ads are named with different names, sometimes the names are the same, but ad content is slightly different, the same ad might be broadcast in different media, and it might have a different name. We watched every single ad and sorted out which ad is which. Then, we aggregated by ad id the expenditures over a month, if the same ad was listed under two different lines.

2. We deflate the ad expenditures by the CPI (January 2000=100)

3. In reporting ad spending, the lowest amount that we can report is $100. Note that we obtain rate card information from media sellers (i.e. TV stations/networks, websites, publications, etc.). Estimated advertising expenditures are assigned to every commercial based on the average 30-second rate for the program. When the commercial is longer or shorter than 30 seconds, the reported dollars are automatically converted in proportion to the number of seconds in the spot. Depending on the length, the reported dollars are adjusted accordingly. Please note that we do not make adjustments for purchased ratings, as we monitor the occurrence level information. Ratings will not affect reported spending. Low prices could be explained by the time of the day (night) or by type of television program (e.g. with low viewership).

4. If a brand is attacking more than one brand with an ad, then we divide the expenditure on that ad by the number of brands attacked to construct how much the brand attacked each one of its competitors.

5. If a brand is attacking prescription drugs, such as Vioxx or non brand specific prescription drugs, then we code it as self-promotion advertising.

6. If a brand is attacking other competitors by mentioning non-brand specific NSAID drug, generic ibuprofen, or other regular OTC pain relief medication, then we code the ad as self-promotion ad.

7. If there was never an attack from one brand to another brand, then we exclude this combination of attacks as a possible attack pair. However, if there ever was at least one attack, then we filled each month of the pair with zero expenditures.

8. If an ad had multiple targets, the ad was assigned equally among them. If an ad had no comparative claims, it was classified as a self-promotion ad. In the data we observe situations when brands made indirect attacks on their competitors. An indirect attack occurs when one brand makes a claim against "all other regular" brands. We code such indirect attacks as self-promotion. We discuss an alternative coding scheme for indirect comparisons in Appendix C.1.

Appendix B: Econometric Model and Identification

B.1 Control Functions and Generalized Residuals

The quality function is written as:

$$Q_j = \alpha \ln\left(A_{jj} + \lambda \sum\nolimits_{k\neq j} A_{jk} - \beta \sum\nolimits_{k\neq j} A_{kj} + \bar{A}_{jj}\right) - \varphi \sum\nolimits_{k\neq j} \ln\left(\bar{A}_{kj} + A_{kj}\right)$$

We now want to show how we can apply the Rivers and Vuong (1988) and the Blundell and Smith (1986) approach when some of the endogenous explanatory variables are left-censored.

We postulate that there exists a vector of instrumental variables Z, and we write:

$$\begin{aligned} s_j &= Z\lambda_1 + u_{1j} \\ \tilde{A}^*_{jk} &= Z\lambda_2 + u_{2j} \\ \tilde{A}^*_{kj} &= Z\lambda_3 + u_{3j} \end{aligned}$$

where $\tilde{A}^*_{jk} = \sum_{k\neq j} A^*_{jk}$, $\tilde{A}^*_{kj} = \sum_{k\neq j} A^*_{kj}$ and A^*_{jk} and A^*_{kj} are the advertising expenditures incurred by the brands. Notice that A^*_{jk} and A^*_{kj} are both left-censored, so that we only observe $A_{jk} = \max\left(A^*_{jk}, 0\right)$ and $A_{kj} = \max\left(A^*_{kj}, 0\right)$. As a result, $\tilde{A}^*_{jk}$ and $\tilde{A}^*_{kj}$ can be (and are in our data) left-censored.

B.1.1 Self-Promotion Equation

Now, write:

$$\xi_j = \rho_1 u_{1j} + \rho_2 u_{2j} + \rho_3 u_{3j} + \upsilon_j$$

So, then:

$$\begin{cases} A^*_{jjt} = -\alpha M s_j - \lambda \sum_{k\neq j} A_{jk} + \beta \sum_{k\neq j} A_{kj} - const \\ -\rho_1 u_{1j} - \rho_2 u_{2j} - \rho_3 u_{3j} - \upsilon_j, \upsilon_j \sim N\left(0, \sigma^2\right) \\ A_{jj} = \max\left(A^*_{jj}, 0\right). \end{cases}$$

And so the issue is how to get u_{1j}, u_{2j}, and u_{3j}. For u_{1j} we just run a simple OLS regression of shares on the IVs, then take the predicted residuals $\hat{u}_{1j}$ and plug them in the regression above.

For u_{2j} and u_{3j} the analysis is more complex, because $\tilde{A}_{jk}$ and $\tilde{A}_{kj}$ are left-censored. We use the notion of generalized residuals as introduced by Gourieroux et al. (1987). Here, the generalized residual $\tilde{u}_{2j}$ is defined as:

$$\tilde{u}_{2j} = E\left[u_{2j}|\tilde{A}_{jk}\right] = \left(\tilde{A}_{jk} - Z_2\lambda_2\right) 1\left[\tilde{A}_{jk} > 0\right] - \sigma_2 \frac{\phi\left(Z_2\lambda_2/\sigma_2\right)}{\Phi\left(-Z_2\lambda_2/\sigma_2\right)} 1\left[\tilde{A}_{jk} = 0\right],$$

where ϕ denotes the normal pdf, Φ denotes the normal cdf, and σ_2 is the standard deviation of u_{2j} that is estimated. The generalized residual $\tilde{u}_{3j}$ is defined in a similar fashion.

In practice, we start by running the Tobit regression

$$\begin{cases} \tilde{A}^*_{jk} = Z_2\lambda_2 + u_{2j}, u_{2j} \sim N\left(0, \sigma_2^2\right) \\ \tilde{A}_{jk} = \max\left(\tilde{A}^*_{jk}, 0\right). \end{cases}$$

We then use the parameter estimates $\hat{\lambda}_2$ and $\hat{\sigma}_2$ to compute the estimated generalized residuals $\widehat{\tilde{u}}_{2j}$. We proceed similarly to determine $\widehat{\tilde{u}}_{3j}$. To estimate α, λ, β, σ, and the constant we run the following Tobit regression:

$$\begin{cases} A^*_{jjt} = -\alpha Ms_j - \lambda \sum_{k \neq j} A_{jk} + \beta \sum_{k \neq j} A_{kj} - const \\ \quad -\rho_1 \hat{u}_{1j} - \rho_2 \widehat{\tilde{u}}_{2j} - \rho_3 \widehat{\tilde{u}}_{3j} - \upsilon_j, \upsilon_j \sim N\left(0, \sigma^2\right) \\ A_{jj} = \max\left(A^*_{jj}, 0\right). \end{cases}$$

Because we are running a regression with generated regressors, we compute the adjusted standard errors with a bootstrap procedure.

B.1.2 Comparative Advertising Equation

As regards endogeneity concerns, the analysis is simpler when we look at the comparative ads first order condition since the only endogenous variables in that equation are the shares of the attacker and of the attacked. So, we can simply use $\hat{u}_{1j}$ and $\hat{u}_{1k}$ and apply the Rivers and Vuong (1988) and the Blundell and Smith (1986) approach again.

In practice, the estimation is made in two steps. First, we run the LHS endogenous variables (here market shares) on all exogenous variables, including those excluded from the second stage relationship. Then, we run the second stage regression (advertising levels here) including the residuals from the first regression as an additional explanatory variable (the "Control Function") to all the second stage explanatory variables. For example, if we want to estimate the parameters of the self-promotion advertising first order condition (ads on sales), we first run shares on generic prices and news shocks, and compute the residuals. Then we run a Tobit where ads are explained by market share, news shocks (if not excluded) and the residuals.

B.2 Standard Errors

For the estimates in Table 3 and 4 we bootstrap the standard errors as follows. We draw 100 independent samples out of the original dataset. On each one of these 100 datasets we rerun the self-promotion and comparative advertising regressions. We store the results and then we take the standard deviation of each coefficient.

For the estimates in Table 5 and 6 we take the 100 samples and use them to compute the damages. Then, for each parameter, we sort them from the largest to the smallest, and we construct the confidence interval using the one at the 5th position and the one at the 95th position.

B.3 Top Brand Dummy

We have investigated various specifications for the fixed effects, and concluded that a specification where there are two fixed effects, one for the top brands (Advil, Aleve, Tylenol), and one for the other brands (Excedrin, Motrin, Bayer) fits our data best. We provide in Figure B1 a graphical description of the relationship between non-comparative advertising and market shares for all brands and months.

FIGURE B1. Relationship between Self-Promotion Ads and Market Shares

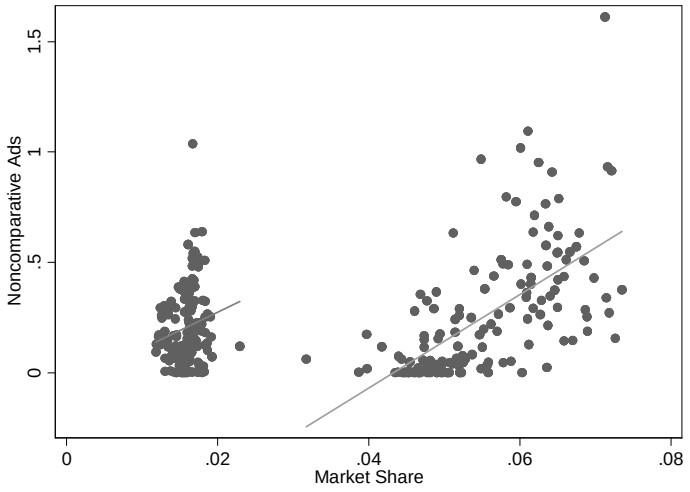


Figure B1 shows that there are two types of brands in the market. Aleve, Advil, and Tylenol (the 'Top Brands') control large market shares compared to Excedrin, Bayer, and Motrin. This is consistent with the reported weighted market share descriptive statistics in Table 1 in the main body of the paper. This observation parallels the economic intuition that 'Top Brands' have a larger advertising base allure which translates into larger inherent quality, $\bar{A}_{jj}$. Additionally, the linear fit between shares and non-comparative advertising has the same slope for the 'Top Brands' and the rest of the brands. We use the evidence from this figure to justify the construction and use of a dummy variable 'Top Brand'.

B.4 Instruments

We assume that generic price can be treated as a proxy for marginal costs in the long run. In the prescription drug markets with multiple generic entrants generic prices usually exhibit a downward price trend and tend to converge to a constant (Grabowski and Vernon,1992). Our instruments would not be valid, if generic prices in our sample exhibited such a downward trend - it would indicate that generic market is still in the process of pricing adjustment. However, one of the most important observations for our identification strategy is that all of the patents for the active ingredients of the OTC analgesics have expired a long time ago. To show that the downward trend in generic prices is not an issue in our industry, we

plot the average price of one serving of a generic product during our sample period. Figure B2 illustrates that the average generic price level does not exhibit a downward trend and fluctuates around the same level. This observation is consistent with the maturity of the generic and branded sub-markets, which have coexisted for decades in the OTC analgesics industry.

FIGURE B2. Variation in Average Generic Prices

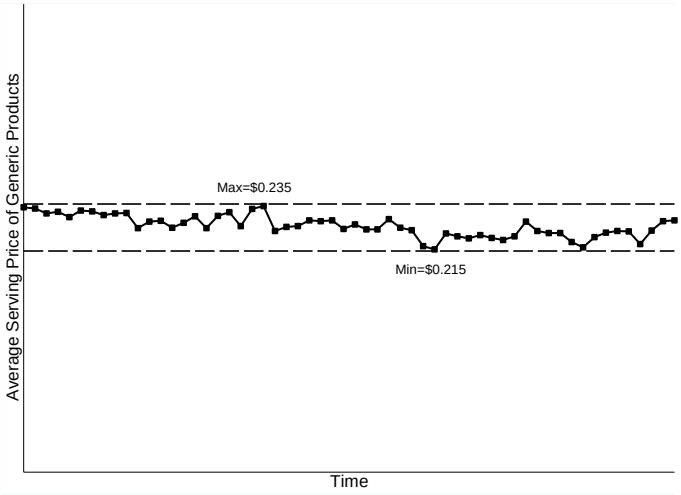


Table B1 reports the first stage results. Exogenous generic prices do a reasonably good job in explaining the variation of the endogenous variables and this result is supported both by the reported R-squared measures and the F tests, which reject the joint hypotheses that the generic price coefficients are equal to zero.

Table B1. Instrumental Variable First Stage Estimates

	Endogenous variables		
	Outgoing Attacks	Incoming Attacks	Market Shares
Generic Acetaminophen	-0.144	0.744	0.232
	(0.697)	(0.892)	(0.510)
Generic Aspirin	-0.032	0.447	0.447
	(0.362)	(0.461)	(0.266)
Generic Ibuprofen	-0.657	-1.242	0.525
	(0.782)	(0.984)	(0.578)
Generic Naproxen Sodium	3.407	3.975	-1.154
	(1.413)	(1.790)	(1.039)
Generic Counterpart	-0.539	-0.254	0.403
	(0.254)	(0.319)	(0.184)
Sum of Generic without Counterpart	-0.321	-0.787	0.018
	(0.251)	(0.319)	(0.183)
Top Brand	0.224	0.479	1.160
	(0.021)	(0.030)	(0.015)
Constant	0.333	0.332	-0.299
	(0.303)	(0.384)	(0.220)
/sigma	0.173	0.207	
	(0.008)	(0.010)	
First stage fit measures			
R-Squared	0.411	0.596	0.946
F(6, 341)	6.27	22.58	40.09
Prob > F	0.0000	0.0000	0.0000
Number of observations	348	348	348

Note: Standard errors in parentheses; The first stage for Outgoing and Incoming attacks is a Tobit estimation and reported R^2 is McKelvey and Zavoina (1975) R^2measure. Market Shares equation is OLS.

Appendix C: Alternative Specifications

C.1 Indirect Attacks

One delicate issue is how to deal with indirect attacks. An indirect attack occurs when one brand, say Tylenol, makes a claim against "all other regular" brands.[3] Because it is not clear how to deal with this type of ad, we consider two solutions. In the main paper we consider the case where indirect attacks should simply be interpreted as self-promotion ads.

[3]Or it could be an attack against NSAIDs (Non Steroidal Anti-Inflammatory drugs), which are all drugs in our sample except those with acetaminophen as an active ingredient.

Here, we consider the case where indirect attacks are equivalent to direct attacks (e.g. Tylenol on Advil), but are divided among all the brands falling within the attacked category. So, for example, when Tylenol makes a claim against "all other regular" brands, each one of the other five brands is being attacked the amount of dollars spent on that advertisement divided by five.

Table C1 presents the results, which should be compared to those in the first two columns in Table 3. Column 1 of Table C1 shows the results when we do not include the Top Brand dummy. Column 2 shows the results when we include that dummy.

The key observation is that the estimates are basically the same as in the first two columns of Table 3. Thus, the coding of indirect attacks is not an empirical concern at all in our empirical study.

TABLE C1. Indirect Attacks

	Coef. (s.e.)	Coef. (s.e.)
α	0.122 (0.027)	0.478 (0.074)
λ	0.804 (0.071)	0.673 (0.073)
β	0.381 (0.065)	0.229 (0.069)
Top Brand dummy		-0.410 (0.080)
Constant	0.132 (0.022)	-0.021 (0.037)
/sigma	0.193 (0.008)	0.186 (0.037)
Log-likelihood	45.302	58.018

31 left-censored observations at PositAdver<=0

317 uncensored observations

C.2 Medical News Shocks

The OTC analgesics market endured several major medical news shocks over the analyzed time period. Following the approach presented by Chintagunta, Jiang, and Jin (2009) we utilized Lexis-Nexis to search over all articles published between 2001 and 2005 on relevant topics. The keywords that we used in our news search consisted of brand names, such as "Aleve," "Tylenol," "Advil," "Vioxx," and the names of their active ingredients, such as "Naproxen" or "Acetaminophen." Then we made searches using generic terms such as "pain killers" or "analgesics." We recorded the article name, source, and date to construct a dataset of news shocks. Multiple articles reporting the same event were assigned to a unique shock ID. Additionally, we checked whether a news shock was associated with any new medical findings that were published in major scientific journals. Finally, we focused only on the events that were reported in a major national newspaper (USA Today, Washington Post, Wall Street Journal, New York Times). After this data cleaning, our news shock dataset included 8 major news shocks between March of 2001 and December of 2005. Table C2 reports the news shocks by their title, date, and the original scientific publication.

After some experimentation, we determined that there are no effects of the news shocks after three months of them happening. We consider two possibilities for the duration of

each news shock in consumer memory. We construct a dummy variable for a short-term shock variant that takes value 1 at time t when the shock occurred, and for the next three months (i.e., t through $t+3$). Then, to check the robustness of our analysis, we construct another variable, which captures the possibility that consumers have a long-term memory. The dummy variable for the long-term shock takes value 1 at time t till the end of the sample period.

TABLE C2. Medical News Shocks

No	News Shock Description	Date	Source
1	Risk of Cardiovascular Events Associated With Selective COX-2 Inhibitors	8/21/2001	Journal of the American Medical Assoc (JAMA); 2001,286:954-959
2	Ibuprofen Interferes with Aspirin	12/20/2001	New England Journal of Medicine, 2001, 345:1809-1817
3	FDA Panel Calls for Stronger Warnings on Aspirin and Related Painkillers	9/21/2002	FDA Public Health Advisory
4	Aspirin Could Reduce Breast Cancer Risk/ NSAIDs Protect Against Alzheimer's	4/8/2003 4/2/2003	JAMA 2004; 291:2433-2440 American Academy Of Neurology
5	Anti-Inflammatory Pain Relievers Inhibit Cardioprotective Benefits of Aspirin	9/9/2003	Circulation, 9/9/2003
6	Vioxx Withdrawn From the Market	9/30/2004	
7	Long Term Naproxen (Aleve) Use may Increased Cardiovascular Risk	12/23/2004	FDA Public Health Advisory
8	Bextra Withdrawn	4/7/2005	

Table C3 reports the model estimates when two types of shock definitions are allowed to affect the base allure, $\bar{A}_{jj}$. In Column 1 we add the variables that measure the occurrence of a news shock using the short term memory definition. The estimates in Table C3 should be compared with the estimates in Table 3 in the paper. With the exception of the estimate of α that increases from 0.432 to 0.513, the results in Column 1 of Table C3 are remarkably similar to those in Column 2 of Table 3, suggesting that adding the short-term memory news shocks as control variables does not change the way the model fits the data. This is consistent with the low values of the F statistic associated with the test that all the coefficients of the news shocks are equal to zero. The results in Column 2 of Table C3 show that adding the long-term memory news shocks as control variables does not change the way the model fits the data either. Furthermore, the estimate of α is the same in Column 3 of Table 3 (specification with generic prices as IVs), suggesting that the instrumental variable approach controls for the endogeneity of s_j to the same extent as adding news shocks does.

Table C3. Effects of News Shocks

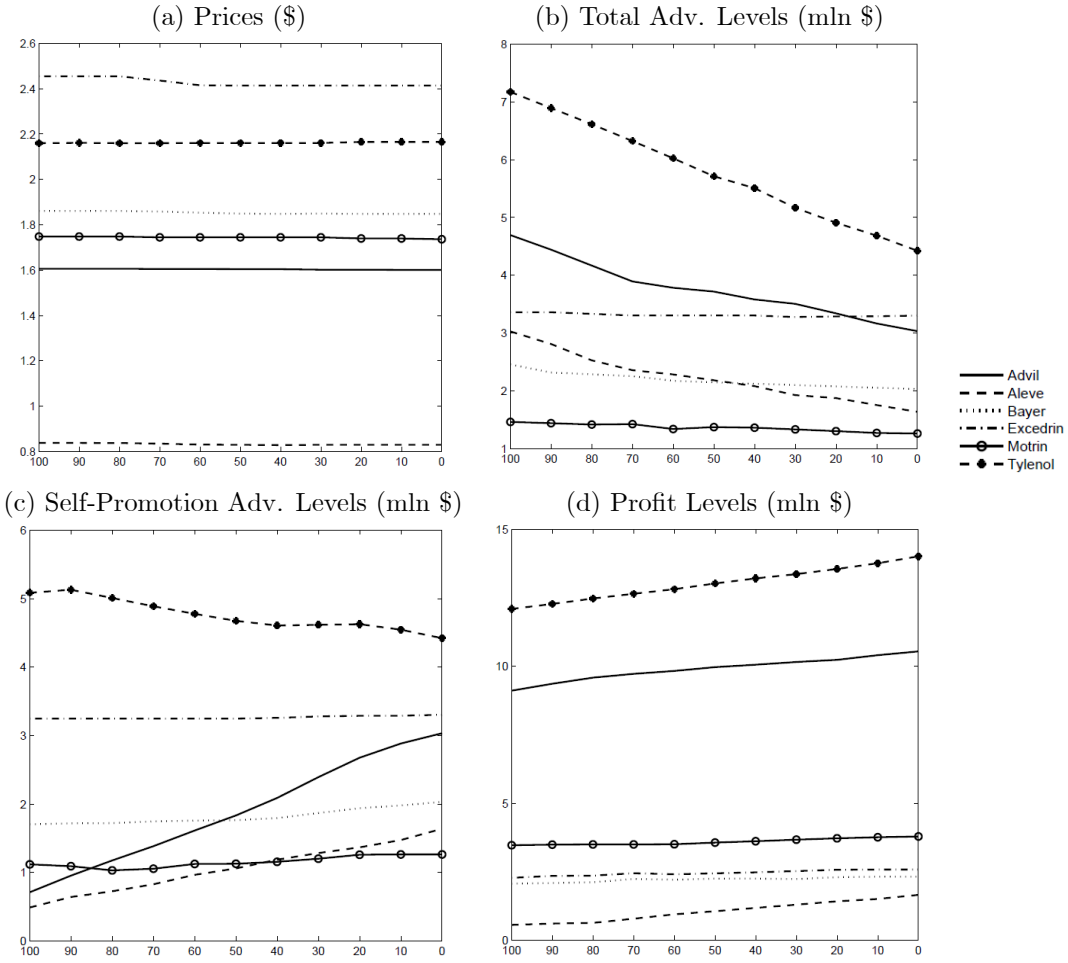
Version	News Shocks Short term (1)	News Shocks Long Term (2)
Alpha	0.513	0.515
	(0.078)	(0.074)
Lambda	0.643	0.631
	(0.073)	(0.071)
Beta	0.251	0.258
	(0.070)	(0.066)
Brand dummy	−0.440	−0.439
	(0.085)	(0.079)
/sigma	0.181	0.175
	(0.007)	(0.007)
Log likelihood	75.089	83.085
F-tests Shocks In	$F(8, 336)$ = 3.70	$F(8, 336)$ = 6.94
Obs	348	348

Appendix D: Counterfactual Analysis

D.1 Additional Information

For the sake of completeness, we provide additional information on the counterfactual exercise presented in the text. Figure D1 presents the counterfactual results for all 6 brands. Similar to the results presented in Figure 1, prices remain largely unchanged for all brands. Total advertising expenditures decline for all brands. In general, the brands that tended to be the biggest targets (Tylenol, Advil, Aleve) exhibit the largest declines in total advertising. Brands that were not targeted (Motrin, Bayer), on the other hand, do not exhibit much change in advertising levels. Self-promotion increases the most for brands that used a lot of comparative advertising (Advil, Aleve) and goes down the most for brands that were targeted the most (Tylenol). Small brands that were never targeted by their competitors (Bayer, Motrin) or had a very small share of attacks (Excedrin) exhibit the smallest changes in self-promotion levels. These results are largely consistent with the fact that brands are using increased levels of advertising as a defensive measure.

Figure D1. Counterfactual results. Median estimates, all brands.



D.2 Robustness Exercises

This section is motivated by the fact that the upper bounds used in Section 7 may be quite loose. We therefore explore another calibration strategy using the symmetry in our theoretical model that cross partials should be equal, $\mathcal{C}_{jk} = \mathcal{C}_{kj}$. First, whenever our estimated diversion ratios allow us to obtain only one of the cross partial using the method described above (in other words, when we have estimated d_{jk} but not d_{kj}) then symmetry directly provides a number for the missing cross partial. Second, when both diversion ratios are estimated, then we impose symmetry by assuming that the two cross partials are equal to the mean of the values obtained from the two estimated diversion ratios. Finally, for those pairs for which we have no diversion ratio, we resort to the upper bound method mentioned above and impose symmetry by again taking the average of the two numbers given by the bounds method.

We now show that the results obtained using this method are to a large extent similar to those obtained by using the bounds method.

Assumption $C_{ij} = C_{ji}$. Recall that

$$C_{ij} = d_{ji}\mathcal{B}_j \tag{1}$$

Therefore, under the symmetry assumption

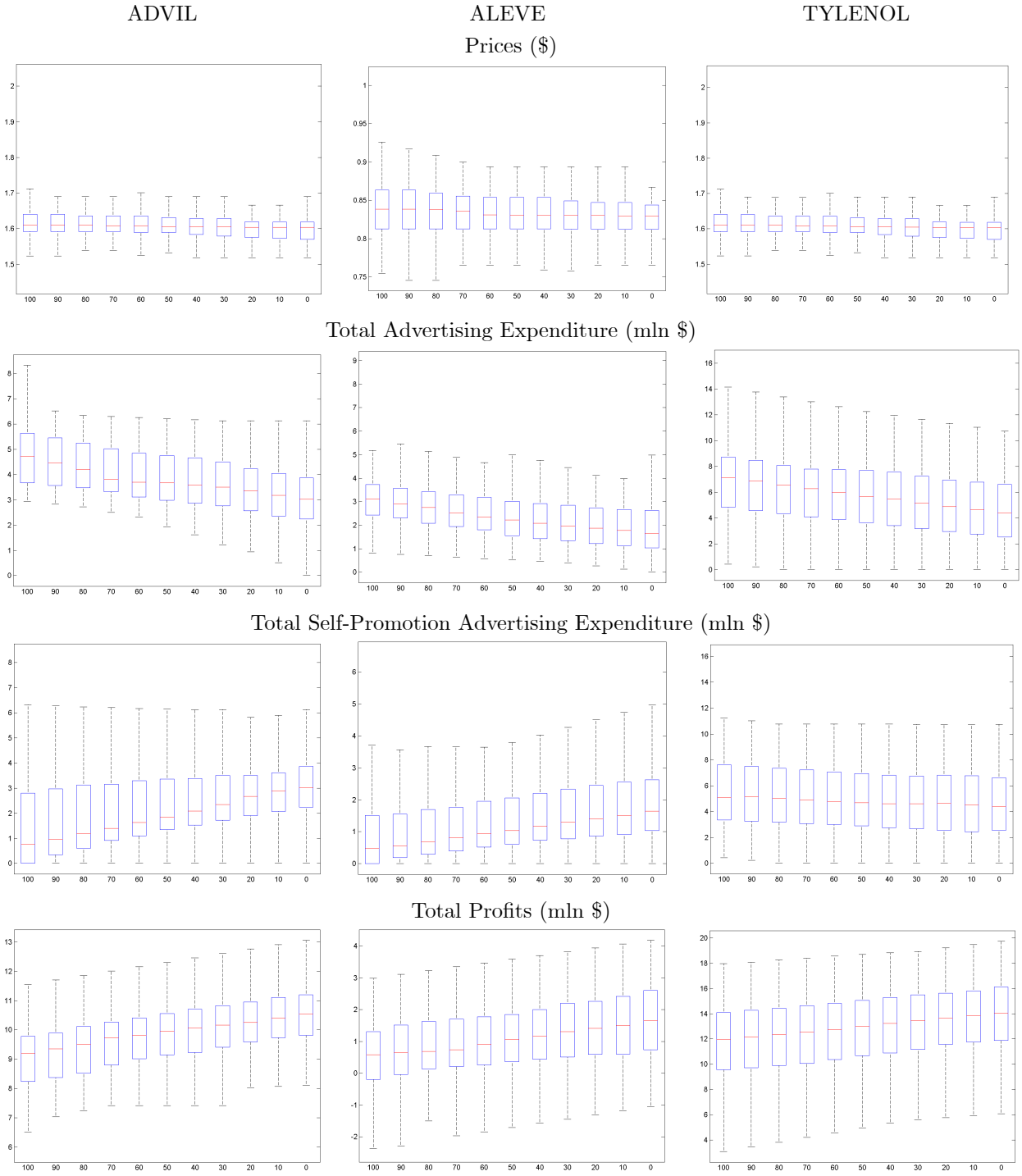
$$d_{ij}\mathcal{B}_i = d_{ji}\mathcal{B}_j \tag{2}$$

therefore, given d_{ji},

$$d_{ij} = d_{ji}\frac{\mathcal{B}_j}{\mathcal{B}_i} \tag{3}$$

We find that twelve diversion ratios can be calculated under the symmetry assumption method. Combined with the diversion ratios estimated in the paper (and accounting for some overlap between the two sets), this means 18 of the 30 necessary diversion ratios are available though this method. Using this approach we can construct the analogue to Figure 1 as follows:

Figure D2. Results of Counterfactual of Banning Comparative Advertising. Symmetry approach.



Notes: The center red line denotes the median; bottom and top of each box correspond to interquantile range (between 25th and 75th percentiles); whiskers extend to the most extreme data points falling within 1.5 times the width of the interquantile range.

The results in Figure D2 are clearly analogous to those presented in Figure 1 in text and Figure D1 above.

References

Blundell, Richard W. and Richard J. Smith (1986), An Exogeneity Test for a Simultaneous Equation Tobit Model with an Application to Labor Supply. Econometrica, 54(3), 679-685.

Chintagunta, Pradeep K., Renna Jiang, and Ginger Zhe Jin (2009), Information, Learning, and Drug Diffusion: the Case of Cox-2 Inhibitors. Quantitative Marketing and Economics, 7(4), 399-443.

Grabowski, Henry G. and John M. Vernon (1992), Brand Loyalty, Entry, and Price Competition in Pharmaceuticals after the 1984 Drug Act. Journal of Law and Economics, 35(2), 331-350.

Rivers, Douglas and Quang H. Vuong (1988), Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models. Journal of Econometrics, 39(3), 347–366.