Supplemental Material for: Bradley T. Shapiro. 2018. "Positive Spillovers and Free Riding in Advertising of Prescription Pharmaceuticals: The Case of Antidepressants." Journal of Political Economy 126(1). DOI: 10.1086/695475.

Appendix A - Orthogonality of Pricing and Detailing Decisions

A.1 Pricing

The regression of television commercials on prices is shown in Table 9. As can be seen, the point estimate is very small and insignificant with respect to both own and cross advertising. The average unit price of a branded drug is about \$3.24 over the course of the sample, and the average DTC per capita for advertising products after September 1999 is \$0.65. The point estimate suggests raising DTC per capita by \$1 is associated with a price decrease of \$0.05. This amount is very small economically. Interestingly, a time trend, a product fixed effect, and a dummy for patent expiration can explain prices with R squared larger than 0.99. Prices seem quite sticky, especially relative to advertising in this market.

Table 1: Predicting Prices with Advertising

VARIABLES	(1) realunitprice
DTC	-0.0478
	(0.0442)
$RivalDTC$	-0.0043
	(0.0053)
time	0.0080***
	(0.0023)
expired	-0.1279
	(0.1153)
Product FEs	X
Observations	1189
R-squared	0.993

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Product clustered standard errors in parentheses

A.2 Detailing

To address the potential omitted-variables bias problem associated with physician detailing, I use two data sources to show physician detailing is uncorrelated with DTC advertising. First, I use national-level detailing information from IMS health and exploit the law change that made DTC advertising feasible for pharmaceutical firms starting in 1999. Using these data, I show no trend break occurs in firm detailing associated with the policy change, although an enormous trend break occurs in DTC, from zero to highly positive. Second, I employ data from ImpactRx that follow a panel of physicians through time, measuring the number of visits and amount of detailing time that the drug spent on them. The data include a panel of 2,134 general practitioners and 167 psychiatrists that can be matched with the DTC and prescribing data. These physicians are not representative of the population of physicians, because they are drawn disproportionately from the 40th percentile and higher in the prescribing distribution. ImpactRx makes this draw because those below the 40th percentile are much less likely ever to be detailed. Whereas the 2,301 physicians make up less than 1% of the population of physicians, they are likely to make up significantly more of the prescription share of the physician population. For general practitioners, the panel begins in 2001, and for psychiatrists, the panel begins in 2002 and includes geographic information. As such, the data are matched with a portion of the data used in the main analysis to test the robustness of the identification to the inclusion of detailing. Detailing is measured in number of detail visits per product per physician per month. With 101 DMAs in the data, each DMA thus has an average of about 23 physicians in this sample.

Using these data, I use two approaches to address the omitted-variables-bias concern. First, I aggregate the local detailing measure to the DMA level and show DTC and local detailing are not correlated conditional on the fixed effects included in the model. Similarly, I aggregate the data to the border-experiment level and show this lack of correlation extends to the border regions. Next, I include the border-experiment-level detailing data in the model for the limited months and experiments that can be matched, and show the exclusion of detailing has no effect on the estimates of DTC effectiveness.

Table 10 presents the results of the national-level analysis using IMS national detailing visit numbers. The regressions are specified using product-specific time trends, product fixed effects, a dummy to indicate the product is in the market, a dummy to indicate the product's patent has expired, and a dummy for "post" September 1999, when the final FDA guidance was released. The coefficient on "post" will identify the average break in trend of detailing at the moment the law changes. The coefficient is small and positive but insignificant. By contrast, using the same specification but with DTC as the dependent variable in the second column, the "post" dummy shows a very large positive and significant effect, as firms move from zero DTC advertising at the national level to a significant amount of DTC advertising. Even at the aggregate national level, this major law change affecting firm strategy appears to have little effect on the path of firm detailing decisions.

Table 2: Did the DTC Law Change Result in a Detailing Trend Break?

VARIABLES	National Details	National DTC
$PostSept99$	0.672	472***
	(0.796)	(124)
Product Time Trend	X	X
Observations	992	992
R-squared	0.932	0.266

*** p<0.01, ** p<0.05, * p<0.1

Standard errors in parentheses. Included in each specification are product fixed effects, product specific time trends, a dummy indicating a product's patent has expired, and a dummy indicating that the product is in the market. Average detailing per month in the sample is 16.807. Average national DTC per month in the sample is 427.75

The next strategy employed to assess the potential for bias is to aggregate the physician-level detailing from ImpactRx to the DMA level. Because DTC advertising is decided at the DMA level and no more local, I want to test if detailing decisions are sensitive to DTC decisions at the DMA level. I regress own local DTC and rival local DTC, measured in dollars per 100 capita on product-detailing visits in the DMA. Similar to the main specification, I include product-quarter and product-DMA fixed effects. I want to know within time period and within DMA if higher DTC for one product is associated with higher detailing. The results of the regression are presented in the first column of Table 11. I find no statistically significant relationship between either own or rival DTC and detailing decisions. This finding is consistent with conversations with managers who said detailing and DTC decisions are made in different divisions of the firm and that coordination is very difficult. Although the estimates are not statistically significant, note that even if they were, the magnitudes of the point estimates are very small. The standard deviation of local DTC campaigns is 0.17. Taken at face value, an increase in own local DTC of one standard deviation is associated with an increase in detailing visits for doctors in this sample in the DMA by about 0.1. Extrapolating this point based on the size of the sample, the point estimate would imply an average of about one additional detailing visit per 230 physicians associated with a one-standard-deviation increase in local DTC. The point estimate on rival advertising is even smaller. This finding provides evidence that widespread strategies to undercut rival DTC with detailing campaigns are not prevalent.

In the second column of Table 11, the same exercise is run at the border-experiment level. The results are broadly similar, though these results should be taken with a grain of salt because each observation is based on far fewer physicians at this narrow level of analysis. Thirty-four of the border areas have physicians on either side of the border, with an average of 3.75 physicians on each side. I find no evidence that detailing efforts change sharply at borders where one side has significantly higher DTC.

Table 3: Are Detailing Campaigns Coordinated with DTC?

VARIABLES	DMA Level Detailing Visits	Border-Experiment Level Detailing Visits
Local DTC	0.662	0.579
	(0.475)	(0.490)
Rival Local DTC	0.184	0.0326
	(0.177)	(0.108)
Observations	22,589	18,622
R-squared	0.764	0.789

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Product-DMA clustered standard errors in parentheses. Column 1 presents DMA-level analysis, whereas column 2 provides border-experiment-level analysis. They contain the same fixed effects as those in the demand models of the same level.

The final strategy used to think about this problem is to incorporate local detailing data directly into the main analysis. Detailing data are aggregated to the product-border-DMA-time level as with the main specification. Because the physician panel in the detailing data is smaller in both number of physicians and in length of time, the matched sample will be substantially smaller and over a shorter time frame than the main specification. As such, T is no longer large enough to ensure the lagged dependent variable Nickel bias is small. The persistence parameters will all be biased downward in these specifications. The main goal of this exercise is to see if the coefficients on DTC are robust to the inclusion of detailing.

The effect of including detailing in this model is demonstrated in Tables 12 and 13 . Both tables include all matched data from 2001 onward. The regressions in Table 12 include the log of detailing visits, whereas the regressions in Table 13 do not. Table 12 presents specifications that both include and exclude interaction terms between DTC and detailing. As we can see, the exclusion of detailing or the interactions of DTC with detailing does not have a significant impact on the effects of DTC. As expected, the persistence parameters are all much smaller than in the main model in Table 5 due to the smaller number of time periods.

These exercises show that over the course of the time period in which detailing data are available in this class, detailing is not significantly correlated with DTC at either the DMA level or the border-experiment level. Further, including

Table 4: Results of Limited Model Including Geographic Detailing from 2001 On

VARIABLES	Category Level	Category Level	Subcategory Level	Subcategory Level	Product Level	Product Level
adstock	0.0484***	0.0474**	0.00345	0.00710	0.0300**	0.0314**
	(0.00638)	(0.00730)	(0.00922)	(0.00968)	(0.0103)	(0.0105)
LogVisits	0.00570	0.00806	-0.00471	0.00029	0.00236	0.00378
	(0.00578)	(0.00806)	(0.00568)	(0.00733)	(0.00358)	(0.00362)
$LogVisits \times adstock$		0.00157		-0.00682		-0.00558
		(0.00357)		(0.00485)		(0.00659)
persistence	0.489***	0.489***	0.124***	0.124***	0.260***	0.260***
	(0.0497)	(0.0496)	(0.0146)	(0.0146)	(0.0153)	(0.0153)
Observations	9,861	9,861	42,860	42,860	77,891	77,891
R-squared	0.960	0.960	0.940	0.940	0.968	0.968

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Level-DMA clustered standard errors in parentheses. Demographic interactions included as in the main analysis but suppressed for expositional clarity.

Table 5: Results of Limited Model without Geographic Detailing

VARIABLES	Category Level	Subcategory Level	Product Level
adstock	0.0491***	0.00301	0.0301***
	(0.00636)	(0.00911)	(0.00103)
persistence	0.489***	0.124***	0.260***
	(0.0497)	(0.0146)	(0.0153)
Observations	9,861	42,860	77,891
R-squared	0.960	0.940	0.968

*** p<0.01, ** p<0.05, * p<0.1

DMA clustered standard errors in parentheses. Demographic interactions included as in the main analysis but suppressed for expositional clarity.

detailing in the main model does not change the main parameters of interest. Although this study makes no claims of identifying the causal effects of detailing, detailing appears to pose a minimal threat to the identification of the causal effect of DTC using the border strategy.

Appendix B - Placebo Test

With a difference-in-differences model, the assumption of parallel trends in the outcome variable absent the treatment is required for a valid estimation. One way to assess the validity of this assumption is through the use of a placebo test. In this case, I will use advertising for over-the-counter (OTC) sleep aids as a placebo treatment. This placebo is ideal for two reasons: first, it varies at the same level as antidepressant advertising at the DMA month. Next, OTC sleep aids need not be prescribed by a physician, so we should not expect a “going to the doctor” effect of advertising to be present in OTC advertising. I will use the same identification strategy, but I will use OTC sleep aid advertising as a treatment. The results are presented in Table 14. None of the coefficients on OTC sleep aid advertising are statistically significant at the 5% level or economically important at any level.

Table B1: Results of Base Model with Placebo

VARIABLES	Category Level	Subcategory Level	Product Level
OTCSLEEPAD	0.000074	0.000471	0.00219*
	(0.000398)	(0.000483)	(0.00113)
persistence	0.684***	0.280***	0.323***
	(0.0300)	(0.0116)	(0.0139)
Observations	22,926	92,295	147,443
R-squared	0.948	0.935	0.960

*** p<0.01, ** p<0.05, * p<0.1

Level-DMA clustered standard errors in parentheses

Supplemental Material for: Bradley T. Shapiro. 2018. "Positive Spillovers and Free Riding in Advertising of Prescription Pharmaceuticals: The Case of Antidepressants." Journal of Political Economy 126(1). DOI: 10.1086/695475.

Appendix C - Alternative Sample Selection

One might worry the effect of advertising in the border-sample counties differs systematically from the non-border counties or that the effect of advertising in rural areas would be much different than the effect of advertising in urban areas. To consider these concerns, I have repeated the analysis with several alternative sample selections.

C.1 The Urban-Rural Divide

C.1.1 Without the Northeast Corridor and Other Urban Areas

It seems the less urban areas show very similar results to the full border sample, as seen in Table 15. The category- and product-level effects are larger, but not significantly different from the full sample of borders.

Table C1: Results of Base Model without Northeast Corridor and Other Urban Areas

VARIABLES	Category Level	Subcategory Level	Product Level
adstock	0.0508***	0.00874	0.0279***
	(0.00825)	(0.00904)	(0.00939)
persistence	0.689***	0.255***	0.313***
	(0.0322)	(0.0129)	(0.0159)
Observations	17,400	67,976	99,101
R-squared	0.946	0.927	0.955

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Level-DMA clustered standard errors in parentheses

C.1.2 Only the Urban Border Counties

The effects of advertising seem a bit smaller in the more urban areas, as shown in Table 16, except at the subcategory level. However, the identifying assumption is more likely to fail in the more urban areas, because the borders are much closer to the central cities than in the more rural borders.

Table C2: Results of Base Model with Only Northeast Corridor and Other Urban Areas

VARIABLES	Category Level	Subcategory Level	Product Level
adstock	0.0348***	0.0191**	0.00707***
	(0.00951)	(0.00940)	(0.0111)
persistence	0.648***	0.387***	0.381***
	(0.0678)	(0.0215)	(0.0284)
Observations	5,775	25,550	41,586
R-squared	0.963	0.956	0.972

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Level-DMA clustered standard errors in parentheses

C.2 The “Anti”-Border Sample

Here, I use all of the counties that are not in the border sample. For this sample, endogeneity should be the most severe. The results of this estimation should reveal the main drivers of the endogeneity problem that necessitates use of the border sample. If the main effects of advertising are much larger in this estimation, we might believe advertising was strongly correlated with unobserved local demand shocks such as weather shocks, employment shocks, or special seminars that are unobserved. These shocks, if positively correlated with advertising decisions, would lead us falsely to conclude the effect of advertising is high. If, on the other hand, the reverse-causality concern–in which firms target an advertising-to-sales ratio based on sales in the previous period–we would expect that in the anti-border sample, the persistence parameter would be much higher. That is, lagged shares would not only be measuring the persistence of prescribing associated with advertising carry-over; they would also be controlling for the rule-of-thumb decision of firms to target DMAs that had large shares in the previous period. Results of the estimation are in Table 17. As can be seen, although the point estimates on advertising do not significantly change from the border sample, they are larger, and the persistence parameters are much (and significantly) larger. This finding suggests the main source of endogeneity is reverse causality: if firms are following a rule-of-thumb advertising-to-sales ratio based on previous-period sales, the lagged-share variable will be biased upward as it controls for the firm’s rule of thumb rather than estimating persistence.

Table C3: Results of Base Model Using the “Anti”-Border Approach

VARIABLES	Category Level	Subcategory Level	Product Level
adstock	0.0508***	0.0049	0.0182***
	(0.00164)	(0.00251)	(0.00439)
persistence	0.751***	0.721***	0.594***
	(0.0259)	(0.0128)	(0.0115)
Observations	8,466	42,181	73,628
R-squared	0.981	0.984	0.984

*** p<0.01, ** p<0.05, * p<0.1

Level-DMA clustered standard errors in parentheses

C.3 Removing the State Borders

As described in the text, if tax rates cause a selection problem of physicians into DMAs, results could be biased. To assess this possibility, I remove from the sample DMA borders that coincide with state borders. Doing so removes roughly one third of the borders in the sample. I re-run the analysis using the set of DMA borders that do not coincide with state borders. Results are available in Table 18. The results are similar and not statistically distinguishable from those in the full sample.

Table C4: Results of Base Model without State-Border Overlap

VARIABLES	Category Level	Subcategory Level	Product Level
adstock	0.0376***	0.0131*	0.0256***
	(0.00764)	(0.00783)	(0.00881)
persistence	0.683***	0.271***	0.327***
	(0.0368)	(0.0133)	(0.0166)
Observations	15,975	65,097	97,720
R-squared	0.981	0.984	0.984

*** p<0.01, ** p<0.05, * p<0.1

Level-DMA clustered standard errors in parentheses

Appendix D - Primary Care Service Areas

As mentioned in section (3.1.1), measurement error could bias my estimates toward zero if patients are going to the doctor in different counties than those in which they watch television advertisements. The Dartmouth Center for Health Policy Research has developed a Primary Care Service Area (PCSA) project that is the first national database of primary care resources for small areas. These areas were defined using Medicare claims data from 1999 and Census data from 2000. The service areas include a ZIP area with one or more primary care providers and any bordering ZIPs where the population largely gets their primary care from those physicians.

This database allows me to ask how many patients travel across DMA borders to seek their primary care. In particular, I can match these data to my prescribing data at the ZIP level. I can then see what percentage of each PCSA falls into a single DMA. By so doing, I find that only about 1% of PCSAs cross DMA borders at all. Those that do cross DMA borders do so only minimally. That is, the DMA holding the majority of a PCSA that crosses a border on average contains 97% of that PCSA. As such, measurement-error bias should be minimal.

Although this conclusion only holds for Medicare beneficiaries and primary care physicians, note that primary care physicians do make up a large portion, on the order of 80% of antidepressant prescriptions.[1] For there to be a large measurement-error issue, non-Medicare recipients must travel larger distances than Medicare recipients to their physicians. The extent to which this is true will lead to measurement error that biases all estimates of advertising effectiveness toward zero.

Appendix E - Selection of Households and Physicians to Counties

E.1 Selection Across the Borders

The possibility that households and physicians that are pre-disposed to being more affected by advertisements systematically select into border counties on the side of the border that receives more advertisements might present a problem.

---

[1] http://psychcentral.com/news/2011/08/08/more-antidepressants-prescribed-by-non-psychiatrists/28423.html.

Addressing this level of selection is difficult. However, I do check for balance in characteristics across the borders. First, I check to see if physicians are more likely to bunch on the "high DTC" side of the border. To do so, I collected data from the Area Resource File from 2001 that contained information on the number of physicians in a particular county. I collected these data for both total number of physicians and non-federal physicians. To determine the "high DTC" side of the border, I totaled the local ads from 2001 for each side of a particular border. If one side saw more advertising in total in 2001, I refer to it as the "high DTC" side. I then ran a t-test to see if the means of the numbers of physicians were the same on either side of the border. The t-test fails to reject that the number of physicians is the same on both sides of the border, with p=0.21 for total doctors and p=0.25 for non-federal doctors.

Next, we might be concerned that the counties on the high-DTC side of the border are systematically higher in income or in population than those on the low-DTC side of the border. A t-test fails to reject that average income is the same across the border, with p=0.47. Similarly, a t-test for population fails to reject that the populations are the same across the border, with p=0.17.

Results for all these tests are presented in Table 19. Note that the market fixed effects in the model allow for different levels in all of these variables as long as the trends are parallel and the effectiveness of advertising is the same across the border. That systematic differences in characteristics do not seem to exist across the border further solidifies that this level of comparison is reasonable.

Table E1: Tests for Selection across Borders

variable	Est.	Std Err.	p-stat	Mean
Physicians	855	715	1.25	0.215
	(136)	(124)		
Non-Fed Physicians	98.8	88.0	1.15	0.252
	(12.9)	(11.3)		
Income	24,162	23,881	0.723	0.471
	(460)	(399)		
Population	1,150,642	979,142	1.37	0.173
	(168,114)	(180,925)		

N = 153 border clusters

Means Above Standard Errors in Parentheses

A further test of the randomness of the borders is to test whether being on the high advertising side of the border predicts observable characteristics of those markets. To assess this, I run a regression with total advertising in the category on the right hand side and various demographic variables on the left hand side. I do this exercise for all variables included as interactions in the demand model and include border-time fixed effects. As such, the comparison being tested is given a particular border pair, whether being on the high side of the border in a particular month predicts the observable characteristics. The findings are reported in Table 20. Only one variable is predicted by category advertising in a statistically significant way, and that is the employment to population ratio, which is on average 0.014 higher for every $1 per 100 capita of additional category advertising. The average category advertising is about $4 per 100 capita and the average employment to population ratio is 0.457. So even though the employment to population is systematically higher on the high advertising side of the border, the difference is very small. Note again that fixed effects in the model do allow for different levels in these variables as long as the trends remain parallel.

Table E2: Tests for Demographic Selection Across Borders

Variable	Est.	Std. Err.	p-value	Mean
$PctBlack$	-0.0056	0.0058	0.335	0.0744
$PctHispanic$	-0.0076	0.0047	0.114	0.0647
$PctAsian$	0.0028	0.0028	0.310	0.0137
$PctUrban$	0.0368	0.0338	0.279	0.596
$PctUninsured$	-0.0043	0.0028	0.127	0.154
$PctOver45$	0.0102	0.0064	0.116	0.378
$PctMale$	0.0019	0.0012	0.121	0.492
$PctEmployment$	0.0141	0.0060	0.022	0.457
$Log(income)$	0.0283	0.0240	0.241	10.087

These regressions reflect estimates from border-area-DMA-month level regressions with the variable listed on the right hand side and category total DTC per 100 capita on the right hand side, including border-time fixed effects

E.2 Selection into the Border Sample

We might also worry the border counties are systematically different from non-border counties in observable ways. This finding would hurt the generalizability of the results. As such, in Table 21, I compare border counties with interior counties in terms of the average number of doctors, the average number of non-federal doctors, the average population, and the average income. These comparisons are at the county level rather than at the border cluster as in the previous comparison, so I have many more observations.

Table E3: Tests for Observable Selection into Border Sample

variable	Border Sample	Non Border Sample	t-stat	p-value
Physicians	297.404	330.548	0.792	0.428
	(864.491)	(1175.56)		
Non-Fed Physicians	34.827	36.932	0.553	0.580
	(81.697)	(99.333)		
Income	24,203	24,413	0.763	0.446
	(6249)	(6201)		
Population	127,310	123,482	0.286	0.774
	(287,262)	(346,498)		

N=664 border, N=2221 non-border

Means above Standard Errors in Parentheses

Appendix F - Framework for Simulating the Size of the Incentive Effects

F.1 Supply

Firm free riding may be an optimal strategy in a game with positive spillovers. To investigate the incentives generated by the demand problem above, we would like to measure the marginal revenues generated by this type of advertising and compare those with the marginal costs. Because advertising has a dynamic effect through carry-over, one must incorporate the dynamics through solving a dynamic programming problem. Note here that the purpose of this exercise is not to predict firm behavior, but to illustrate the size of the incentive effects of advertising spillovers. To do so, I will assume a very stylized model in which firms optimize profits only over advertising and are forced to hold pricing and detailing fixed.

I assume firms play a simultaneous game, choosing advertising each period while taking into account expectations of rival behavior and the dynamic effects of advertising. This approach enables me to analyze the magnitude of potential under-provision levels on average generated by positive spillovers in a very stylized game in which only DTC advertising is moving.

F.1.1 The Firm's Problem

A forward-looking firm maximizes a discounted stream of future profits with respect to advertising. Suppose advertising has a constant marginal cost k_{jm}, the market size is μ_m , prices are p_j, marginal production costs are mc_j, and the discount rate is β. Further, suppose the exogenously given set of products that a firm f has in the market is denoted by Φ_f, and the full set of products in the market is denoted by $\cup_f \Phi_f$.

Per-period profit for the firm will be a function of advertising stock at the product, subcategory, and category levels $\mathbf{A_{mt}}=\{A_{jmt}, A_{nmt}, A_{lmt}\}$, which is a function of the vector of current advertising for all products $\mathbf{a}_{mt}$, advertising stocks in the previous period $\mathbf{A}_{m,t-1}$, and persistence parameters $\lambda_j, \lambda_n,$ and λ_I. Per-period profit is also a function other variables X, a constant marginal cost of advertising k_{jmt}, and product-market-time-specific iid demand shocks ξ:

$$\pi_{fmt} = \sum_{j\in\Phi_f} (p_{jt} - mc_{jt})\mu_{mt} s_{jmt}(\mathbf{A}_{mt}, X, \xi) - k_{jmt} a_{jmt}.$$

The firm's problem is to maximize the stream of future profits for all products in its portfolio. Advertising is set prior to the realization of any demand shocks, so only expected profits matter for firm choices. Expected per-period profits are

$$\pi_{fmt} = \sum_{j\in\Phi_f} \int (p_{jt} - mc_{jt})\mu_{mt} s_{jmt}(\mathbf{A}_{mt}, X, \xi) p(\xi) d\xi - k_{jmt} a_{jmt}. \tag{1}$$

The timing of the game is as follows. At the start of each period, the state of the market $\mathbf{g_{mt}}$ is revealed to all firms. Based on that state, firms will make advertising decisions $\sigma_{jm}(\mathbf{g_{mt}}) = a_{jmt}$. The state variables could include anything in the past history of the game, but I will restrict attention to games in which payoff-relevant state variables are the only ones that matter. Given the demand system provided above, state variables will include only the current advertising stock $\mathbf{A_{mt}}=\{A_{jt}, A_{nmt}, A_{lmt}\}$. After the state variables are observed, the firm makes its advertising decision, the current period demand shocks are realized, and the firm collects its profits for the period.

The strategy profile $\sigma = (\sigma_1, ..., \sigma_J)$ contains the decision rules of the firms. Firm f's expected present discounted value of all profits given the current state $\mathbf{g_{mt}}$ and the full strategy profile σ is

$$V_{fm}(\mathbf{g_{mt}}|\sigma) = \mathbb{E}[\sum_{j\in\Phi_f}^{\infty}\sum_{\tau=t}^{\infty}\beta^{\tau-t}\pi_f(\mathbf{g_{ms}},\,\sigma_f(\mathbf{g_{ms}}))|\mathbf{g_{mt}}]. \tag{2}$$

Firm f chooses a sequence of advertising levels σ_f , which are state dependent, to maximize $V_{fm}(\mathbf{g_{mt}}|\sigma)$. I will assume firms form expectations regarding future states and thus future strategies given current- period state variables. As such, firm f will have a *Markov strategy*, $\sigma_{fm} : \mathbf{g} \rightarrow \sigma_f(\mathbf{g}) = (a_{jm})_{\forall j\in\Phi_f}$. To assess the strategies of competitors, firm f makes an assumption about σ_{-f} and chooses σ_f.

The equilibrium concept used will be *Markov perfect equilibrium.* A *Markov perfect equilibrium* is a set of strategies that form a subgame perfect equilibrium whereby strategies may only be conditioned on payoff-relevant state variables and on the current state of the game. In particular,

$$\sigma_f^* \in argmax\{V_{fm}(\mathbf{g_{mt}}|\sigma)\} \tag{3}$$

for all states $\mathbf{g_{mt}}$, firms f, and actions a_{jm}. That is, each firm will maximize the discounted sum of expected profits given beliefs about competitors' behavior, and each firm's beliefs are mutually consistent in equilibrium. Using MPE makes the allowable strategy space relatively simple. I will further restrict attention to pure strategies in advertising. Firms will use the Bellman equation to solve the dynamic programming problem:

$$V_j(\mathbf{g}|\sigma) = \underset{a\,\in\,\mathbb{R}^+}{sup}\{\pi_j(\mathbf{g},\,a,\,\sigma_{-j}(\mathbf{g})) + \beta\mathbb{E}V(\mathbf{g}'|\sigma)\}. \tag{4}$$

F.2 MPE Simulation

I solve the Markov perfect equilibrium (MPE) of the advertising game using a policy iteration algorithm, similar to Dube et al. 2005. I take an initial guess of the strategy profile $\sigma^0 = (\sigma_1^0, ..., \sigma_J^0)$, and then take the following steps:

1. For the strategy profile σ^n, calculate the value functions for each of the J firms, V_j^n. The value functions are defined by the Bellman equation (31). In this step, the maximization on the right-hand side is not carried out, but the current guess of $\sigma^n$is plugged in.
2. For all $n > 0$, check whether the value functions and policy functions satisfy predetermined convergence criteria, $||V_j^n - V_j^{n-1}|| < \varepsilon_V$ and $||\sigma_j^n - \sigma_j^{n-1}|| < \varepsilon_\sigma$. If so, stop the process.
3. Update each firm's strategy using the Bellman equation (31). Do this by carrying out the optimization on the right hand side. Denote the resulting policies and value functions by $\sigma^{n+1}$and V^{n+1}. Return to step 1.

The value functions are represented on a grid. As in the supply simulation, two firms are present, and two firm advertising stocks and a category advertising stock exist, summing to three state variables. The grid is 5x5x5. Outside of the grid, the value function is determined using bilinear interpolation.

For the two-firm cooperative counterfactual, the MPE simulation turned into basically a monopoly-game simulation whereby both firms' profits are considered and no other firms advertise. For the full cooperative counterfactual, computation is the same, only the advertising firm considers the profits of all products on the market.

Using the derived policies in each of these computations, the equilibria are simulated forward given the market initial conditions to generate the figures.

Available at Request - LDV Simulation

Because of the concern that a lagged dependent variable (LDV) will have a biased estimate with fixed effects in a panel model due to correlation between the lagged dependent variable and the error term, the purpose of this appendix is to simulate the size of such a bias. The theoretical size of the bias is computed in the text, but that size only holds as $N \to \infty$. The finite-sample simulation will assess how good of an approximation that is. In particular, I assume the data-generating process is

- $\lambda_{category} = 0.68$
- $\alpha_{category} = 0.0494$
- $\lambda_{subcategory} = 0.28$
- $\alpha_{subcategory} = 0.00662$
- $\lambda_{product} = 0.324$
- $\alpha_{product} = 0.025$,

with product-border-market, product-border-time, subcategory-border-market, subcategory-border-time, category-border-market, and category-border-time fixed effects drawn from random normal distributions initially, all mean zero and the market effects with standard deviation of 1 and time effects standard deviations of 0.5.

Using 500 draws of epsilons with standard deviations that match the residual standard deviations of product, subcategory, and category dependent variables, I simulate data, and estimate the parameters using the same methods described in the paper. I then compute the difference between the estimated parameters and the true values that I set, $\hat{\lambda}_l - \lambda_l$ and $\hat{\alpha}_l - \alpha_l$. The results of the simulation are as follows:

Table F1: Simulated LDV Bias

Variable	Mean	SD
$(\hat{\lambda}_{category} - \lambda_{category})$	-0.0208	0.00258
$(\hat{\alpha}_{category} - \alpha_{category})$	-0.000790	0.00598
$(\hat{\lambda}_{subcategory} - \lambda_{subcategory})$	-0.0401	0.00171
$(\hat{\alpha}_{subcategory} - \alpha_{subcategory})$	0.00132	0.00933
$(\hat{\lambda}_{product} - \lambda_{product})$	-0.0223	0.000911
$(\hat{\alpha}_{product} - \alpha_{product})$	0.00009	0.00450