Private Labels, Famous Brands, and Heterogeneous Households: Can High Ad Spending be Justified and are Households' Advertising Elasticities Stable Across Products?

Michael Shashoua, Jeremy Fox, and Stefan Hoderlein

This Version: November 20, 2018

Click here for the most recent version

Abstract:
This paper estimates the correlation of household advertising elasticities across various brands of chocolate and laundry detergent. We leverage a unique data set consisting of matched ad views and household purchases across a two year period. This data allows us to address endogeneity issues such as the correlation between ad exposure and household ad elasticities and the distinction between unobserved heterogeneity and state dependence. We extend the dynamic panel methods of Arellano and Bond (1991) to allow time varying random coefficients that can be correlated with regressors (advertising exposure) and correlated across equations in a SUR (seemingly unrelated regressions) system. We also address two specific puzzles. The first involves the high advertising spending in these industries, which we estimate to be well above the Dorfman-Steiner level of optimal spending. The second looks at the strength of the private labels (store brand) and why consumers purchase them despite their lack of television advertising.

1 Introduction

In the past 8 years, global advertising expenditures have grown between 4-7% year-on-year, with expenditures projected to exceed $550 billion in 2018. The proliferation of online platforms such as Facebook, YouTube, Twitter, and Instagram has fueled this explosion of advertising, and companies need to determine 1) where to direct their dollars and 2) how much to spend on advertising. Advertising should be directed to platforms which promote the biggest increase in sales–those platforms that have the highest advertising elasticity of demand. In terms of how much to spend, the ratio of advertising expenditure to revenue varies greatly across industries. For example, the beverage industry has a 3.9% ratio, while hospitals are at 0.3%, and grocery stores are at 0.8%. Even within an industry, competitors may behave very differently; Volvo has a 5.3% ad expenditures to sales, while Ford is at 2.9%, Volkswagen is at 2%, and Toyota is at 1.6%.

Dorfman and Steiner (1954) derive the optimal (profit maximizing) level of advertising for a static firm that owns a single brand and chooses both price and advertising. Their first order conditions imply

$$\frac{\text{advertising \$}}{\text{revenue \$}} = -\frac{\text{advertising elasticity}}{\text{price elasticity}}$$

or that the ratio of advertising to sales equals the negative ratio of advertising elasticity to price elasticity. Intuitively, this states that as advertising becomes more effective, more money should go towards advertising. Conversely, as consumers are more responsive to price changes, advertising expenditure should decrease as consumers cannot be convinced to pay more for goods that advertise more.

One puzzle in the marketing world is the relatively high ratio of advertising expenditure to revenue we see in certain consumer goods. Laundry detergent, for example, has a ratio of over 10%. For a benchmark own price elasticity of -2 for a branded consumer packaged good such as laundry detergent, the advertising elasticity of demand should be around 0.2 to justify the 10% advertising to sales ratio observed in the data. Similarly, the chocolate industry needs an advertising elasticity of demand around 0.1 to justify the 5% advertising to sales ratio observed in practice.

We estimate the advertising elasticity using German household panel data that matches television advertisement exposure and household purchases in the laundry detergent and chocolate category. Chocolate and detergent are two salient industries in Germany, as the 2017 top two advertising companies in Germany were Proctor & Gamble and Ferrero confectionery, which account for more advertising expenditure than the next 6 leading advertisers combined. This type of matched panel data on ad exposure and purchases is rare in the non-internet academic literature on advertising effectiveness. The household panel data cover nearly 4,000 households over a 24 month period and focus on 9 leading detergent and chocolate brands in Germany.

We analyze the top two brands of each industry and a private label brand for each industry. We estimate a model of the household's decision to purchase this focal brand given ad exposure and purchase history. The main outputs we consider are various advertising elasticities. We investigate whether the advertising elasticity for the focal brand is around 0.2 for detergent, the level that explains the 10% advertising to sales ratio, and .1 for chocolate, the level that explains the 5% advertising to sales ratio.

A related puzzle is the large market share of private label, even when an industry has very strong brands. This occurs despite the low promotion and advertising of private labels compared to brand name goods. In general, private-label sales are strongest in commodity-driven, high-purchase categories with little product differentiation. In Germany, nearly 1 out of every 3 dollars spent on consumer packaged goods is for private label. We explore the purchase behavior of these consumers that buy private label and the correlation between private label demand and brand name ad elasticity, both between and within product categories.

A naive investigation using household panel data of this sort to estimate advertising elasticities suffers from at least four sources of endogeneity. 1) Brand-Level advertising may be correlated with brand-specific, aggregate demand shocks. Think about a brand choosing advertising campaigns to coincide with upswings (or downswings) in the fortunes of the brand. 2) Household ad exposure may be correlated with household persistent tastes for a brand. Think about a detergent firm targeting ads to housewives by showing ads during daytime soap operas. 3) Time-specific household demand shocks may be correlated with time-specific household ad exposures. For example, households on vacation are not making any purchases nor are they watching any ads. Such pairs of zero ad exposure and zero purchase will increase the correlation between ad exposure and purchase behavior in the data and bias the estimates of the advertising elasticity upwards. 4) Past purchase history of a household should interact with ads in the purchase decision. For example, laundry detergent is a storable good and those with stockpiled detergent may not respond to ads because they have no need for more detergent. This point involves the importance of distinguishing between unobserved heterogeneity and state dependence.

Our model and unique data combine to form an identification strategy that addresses the four listed endogeneity problems. We estimate a random coefficients model, which allows for flexible patterns of correlation across households, brands, and product categories. We specifically tackle the endogeneity challenges by introducing correlated random coefficients which contain an AR(1) error, and allowing correlation between the random coefficients and the measured variables.

The results show that individual heterogeneity must be taken into account to rationalize the advertising elasticity, although a large portion of the population is not responsive to advertising. This explains the hyper-targeting we see in platforms today, where marketers can use Facebook to target as specifically as "College Educated Males who have been married for less than 3 months who express an interest in coffee and triathlons and who are friends of people that use my service." The confidence intervals show that targeting the most responsive subset yields an advertising elasticity well above .2; however, the average consumer is not responsive to ads. Furthermore, access to panel data across multiple industries allows the inference of correlated responsiveness/stability to advertisements by a household in each industry and for each brand. We see that household preferences for private label are stable across products, and ad responsiveness is positively correlated for chocolate products and negatively correlated for detergent products.

2 Literature Review

In a correlated random coefficient model, the main parameter of interest is generally the average partial effect (APE). However, when the average partial effect is not a constant, that is when the effect of X on the dependent variable Y is heterogeneous, 2SLS is generally not consistent for the APE (Imbens and Angrist 1994). Heckman and Vytlacil (1998) and Wooldridge (2003) show that the 2SLS estimand is consistent for the APE if the effect of the instruments Z on the endogenous variables X are homogeneous. This assumption is particularly uncomfortable in the dynamic panel setting, where lagged dependent variables are among the endogenous variables and instruments include lags of the covariates. In other words, it requires in the first stage $x_{t-\ell}$ to homogeneously affect $y_{t-\ell}$, but x_t can heterogeneously affect y_t in the second stage.

Flores, Heckman, Meghir, and Vytlacil (2008) instead consider a polynomial expression in the second stage with the requirement that the first stage has a function $h(\cdot)$ that is strictly increasing in a scalar unobservable V such that $X = h(Z, V)$. Flores et. al. (2008) show that if both X and Z are continuous the APE can be identified. Masten and Torgovitsky (2016) explain that the APE can be identified even if Z has binary or discrete support. Taken together, Flores et. al. (2008) and Masten and Torgovitsky present an alternative to 2SLS which allow for heterogeneity in the causal effect of Z on X and allows for binary or discrete instruments Z.

In terms of advertising, Bagwell (2007) summarizes the theoretical and empirical work in the field. The common pitfall is accounting for endogeneity in the interpretation of results. For example, are firms with high sales successful on account of their advertising expenditure, or would they also be successful in the absence of advertising. Tellis (1988) uses individual data on sales and television advertising exposure for toilet tissue paper to estimate a model of sales as a function of past advertising exposure. More recently, Ackerberg (2001) uses similar panel data on yogurt to regress sales on television advertising interacted with an individual's purchase history.

A recent literature on internet ad exposure using experimental variation and before/after within-household comparisons (Lewis and Reiley, 2014; Golden, 2014; Blake, Nosko, and Tadelis 2015). This literature tends to find modest effects of internet advertising on sales, although the expenditure levels in the experiments are far below the levels spent by, say, detergent firms on television advertising.

Another literature has used several empirical strategies to isolate plausibly independent variation in aggregate advertising using television media markets. Dubé and Manchanda (2005) and Gordon and Hartmann (2013) exploit ad price variation across media markets. Shapiro (2018) exploits a discontinuity in ads seen on the borders of media markets. Hartmann and Klapper (2017) and Smith, Stephens-Davidowitz and Varian (2017) exploit variation in the identity of the football teams playing in the Super Bowl, which affects viewership depending on the distribution of fans for those teams across markets.

This data we use was also applied in a working paper by Nagel (2010), in which he estimates a discrete choice logit to account for local market advertising. Nagel's work focuses on price endogeneity, while our paper abstracts away from pricing and focuses on advertising decisions and response to advertising. Our paper controls for multiple sources of advertising endogeneity while incorporating time varying random coefficients which are correlated with regressors.

3 Industry Background and Data

3.1 German Media Market

Germany has the largest TV market in Europe, with over 95% of German household's owning at least one television receiver. Germany also has the largest population in the European Union (81 Million) and the highest GDP. The television market can be broken down into three main categories: 1) Public Channels (Government Owned), 2) Private Channels (Free TV), and 3) Pay-TV channels. There are over 300 channels licensed in Germany, but in our data we observed 21 distinct channels.

While the most popular public channels and private channels have similar viewership levels, the public channels account for a very small percentage of the ads viewed. German law does not allow advertisements on its public channels at certain times (for example, after 8pm), and until 2010 the public TV channels could not run "product placement" ads. Table (1) shows the viewership and ad revenue for leading TV channels in Germany in 2006.

The data is from 2004 to 2006, well before internet advertising grew to the behemoth it is today. The ads in our data are evenly distributed at 10, 15, 20, 25, or 30 seconds. Advertising spots are generally sold as either "upfronts," in which television networks sell television airtime several months before the new television season begins, or can be sold as individual slots when there are important TV events.

Pricing for advertising is channel dependent. Some TV stations perform linear pricing, in that they

Table 1: Market Share and Ad Revenue Market Share of Leading German TV Stations, 2006

Television Channel	Type	2006 Market Share	2006 Ad Expenditure Market Share
ARD	Public	14.2%	2.9%
ZDF	Public	13.7%	2.1%
RTL	Private (Free-TV)	12.8%	27.4 %
SAT1	Private (Free-TV)	9.8%	19.6%
PRO7	Private (Free-TV)	6.6%	16.7%
Other Channels		42.9%	31.3%

Source: Statista

charge per second, while others have disproportionate rate structures. Channels often require a minimum booking, such as 120 seconds per week and can sell specific slots. For example, news programming can sell advertising space between the weather and traffic segment. Similarly, networks which show movies can sell split-screen space during the final credits. The average CPM (cost per 1000 views) in Germany for all networks from 2004 to 2006 is 9.5-11€.

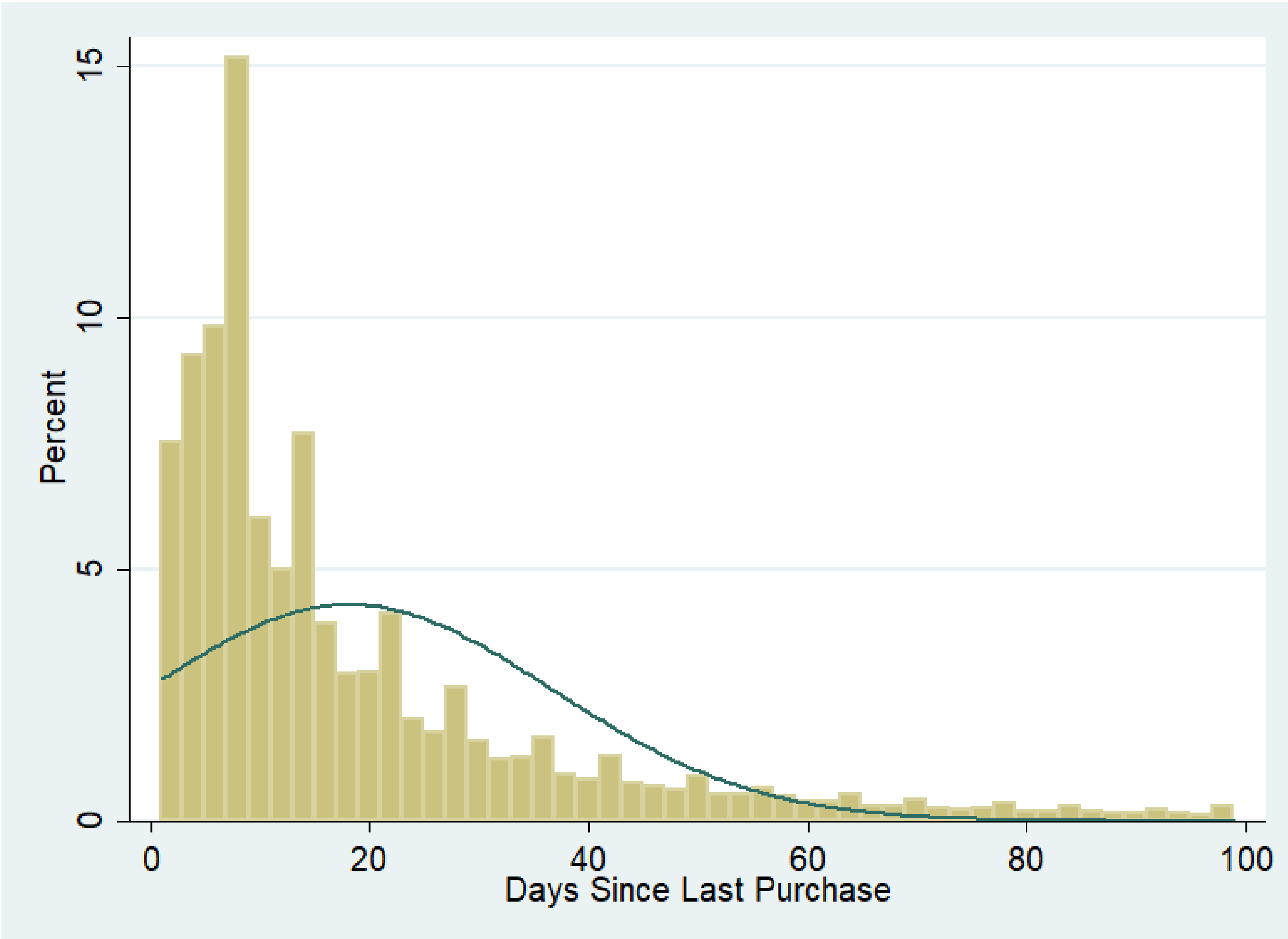
3.2 German Chocolate Industry

Chocolate confectionery is a very large business in Germany. It has the highest retail value in Western Europe and accounts for over $8 billion in sales per year since 2004. The largest company by market share is Ferrero, followed by Lindt, Mondelez, Mars, and private-labels. Each company sells chocolate under multiple brands, and stores also maintain multiple types of private-label brands. A unique feature of the chocolate industry is the large amount of purchases which occur around holidays; seasonal chocolate accounts for a large share of total sales, nearly 15%.

In terms of distribution, 93% of chocolate is sold though grocery retailers, with the lions share of that divided between discounters, supermarkets, and hypermarkets. We describe these different types of stores in more detail below.

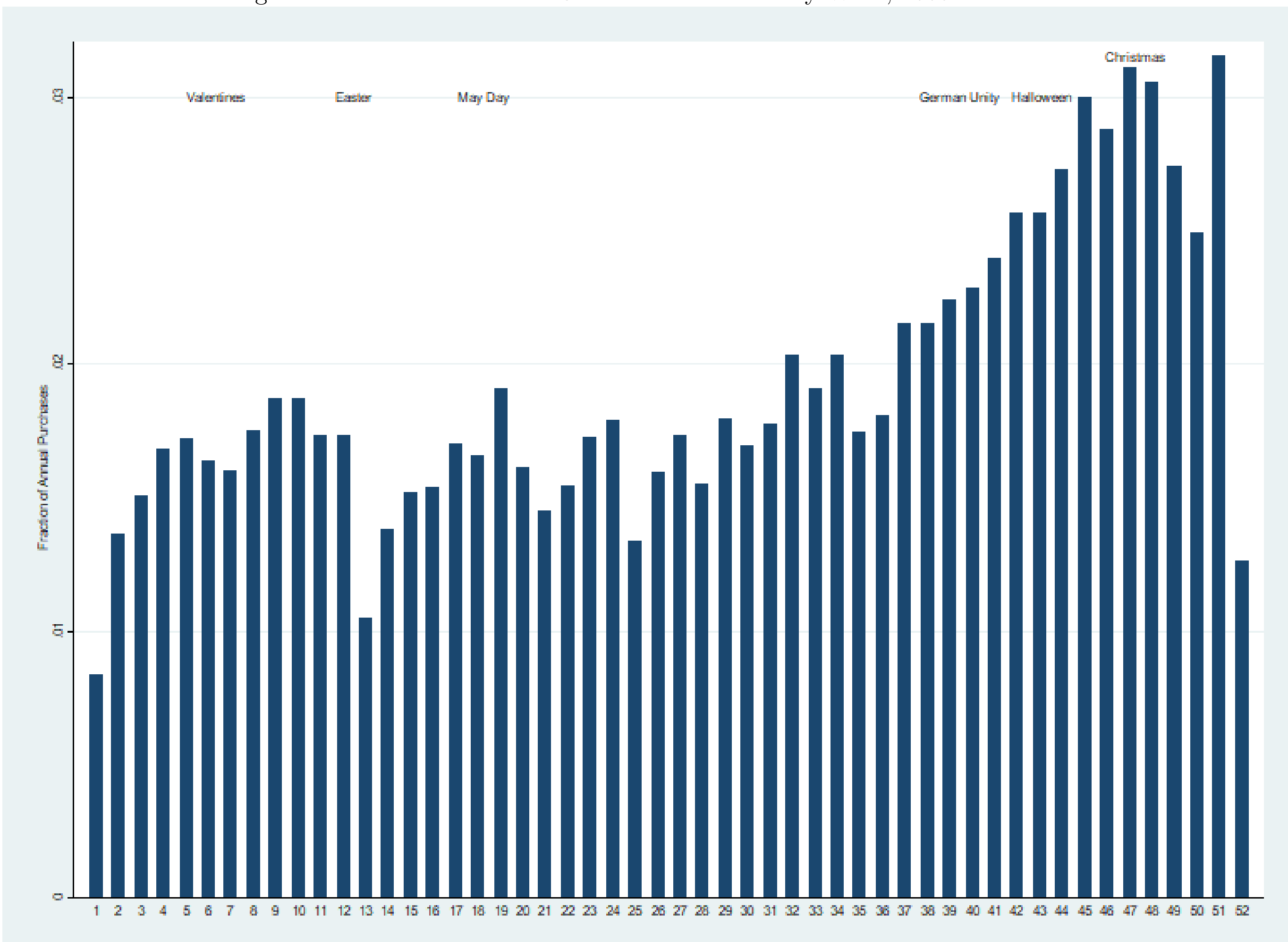
On average, a household purchases chocolate every 18 days, which is twice as frequent as detergent purchases. In figure (1) , we see a rather long tail of some households purchasing chocolate every few months, but nearly 40% of purchases are made within a week of the previous purchase. Furthermore, 37% of the purchases we observe in our data are private label. Figure 1 shows a large heterogeneity in purchase behavior, which our model must account for.

Figure 1: Histogram of Days since last Chocolate Purchase



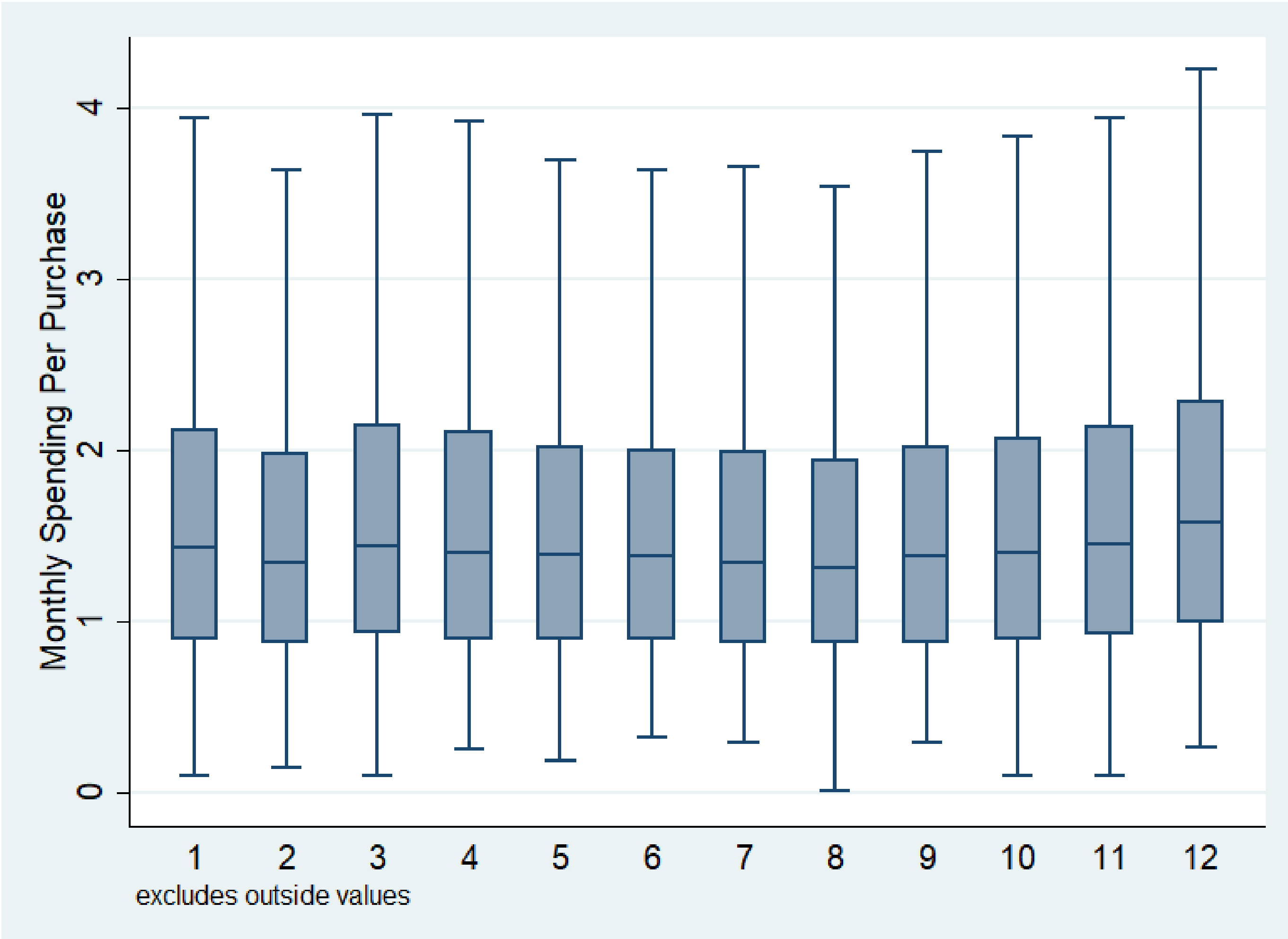
If we focus on 2005 and look at the weekly purchases of chocolate, figure 3.2 shows that there are distinct peaks around Valentines, Easter, May 1st, and towards the end of the year with Halloween and Christmas. The majority of the purchases occur during the holiday season, which follows the stylized fact that a large percentage of chocolate sales are seasonal. Figure (2) suggests that brand-specific time fixed effects are necessary as well.

Figure 2: Fraction of Annual Chocolate Purchases by Week, 2005



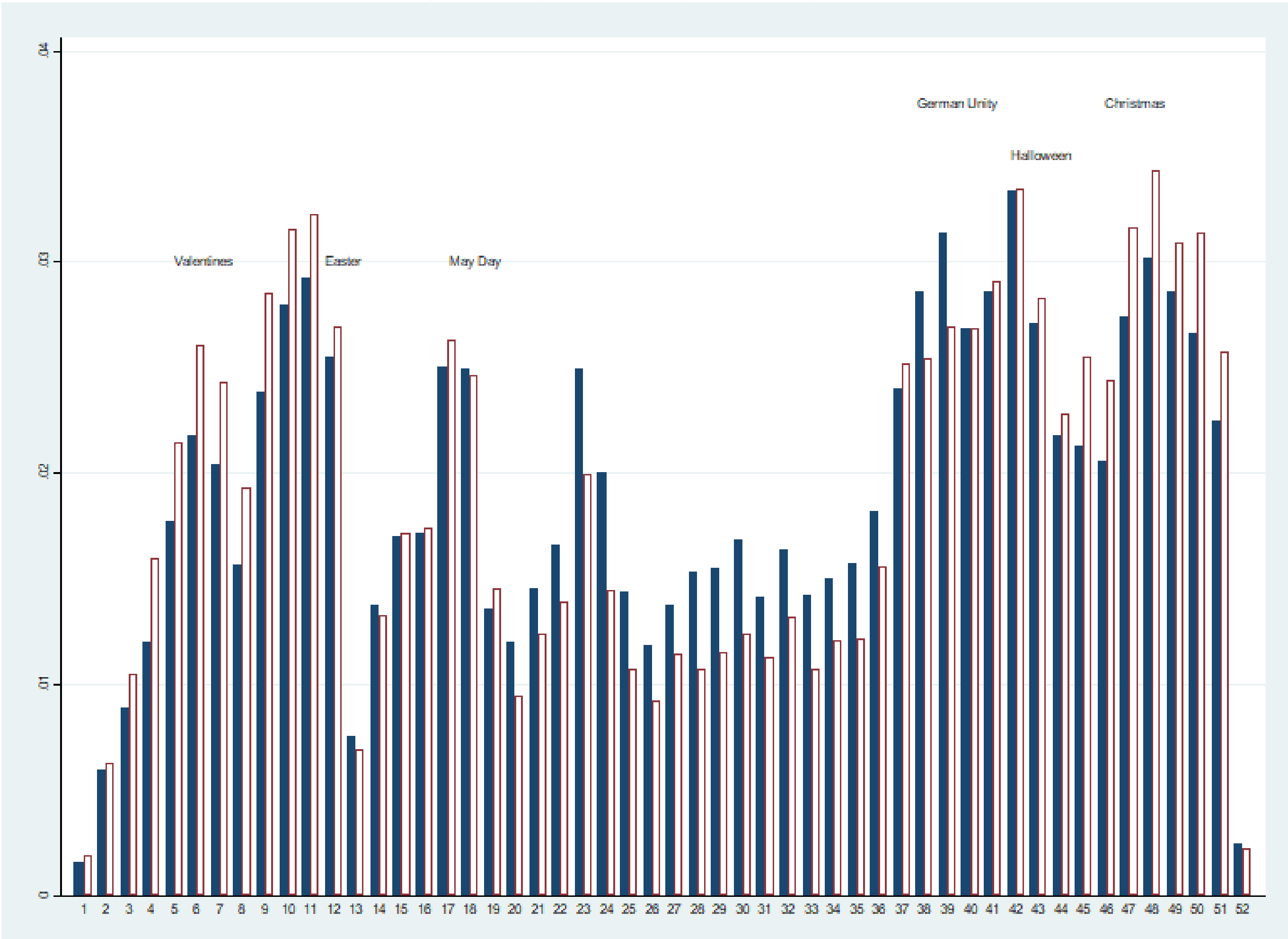
If we look at the overall monthly spending per purchase in figure (3) , the means are pretty similar across months, around 1.5 euros. However, there is a slight increase in spending per purchase in December, along with many more purchases. Figures (2) and (3) suggest that the season growth is not caused by the same households purchasing more chocolate, but rather more households making purchases in general.

Figure 3: Boxplot of Average Monthly Chocolate Spending per Purchase by Month, 2005



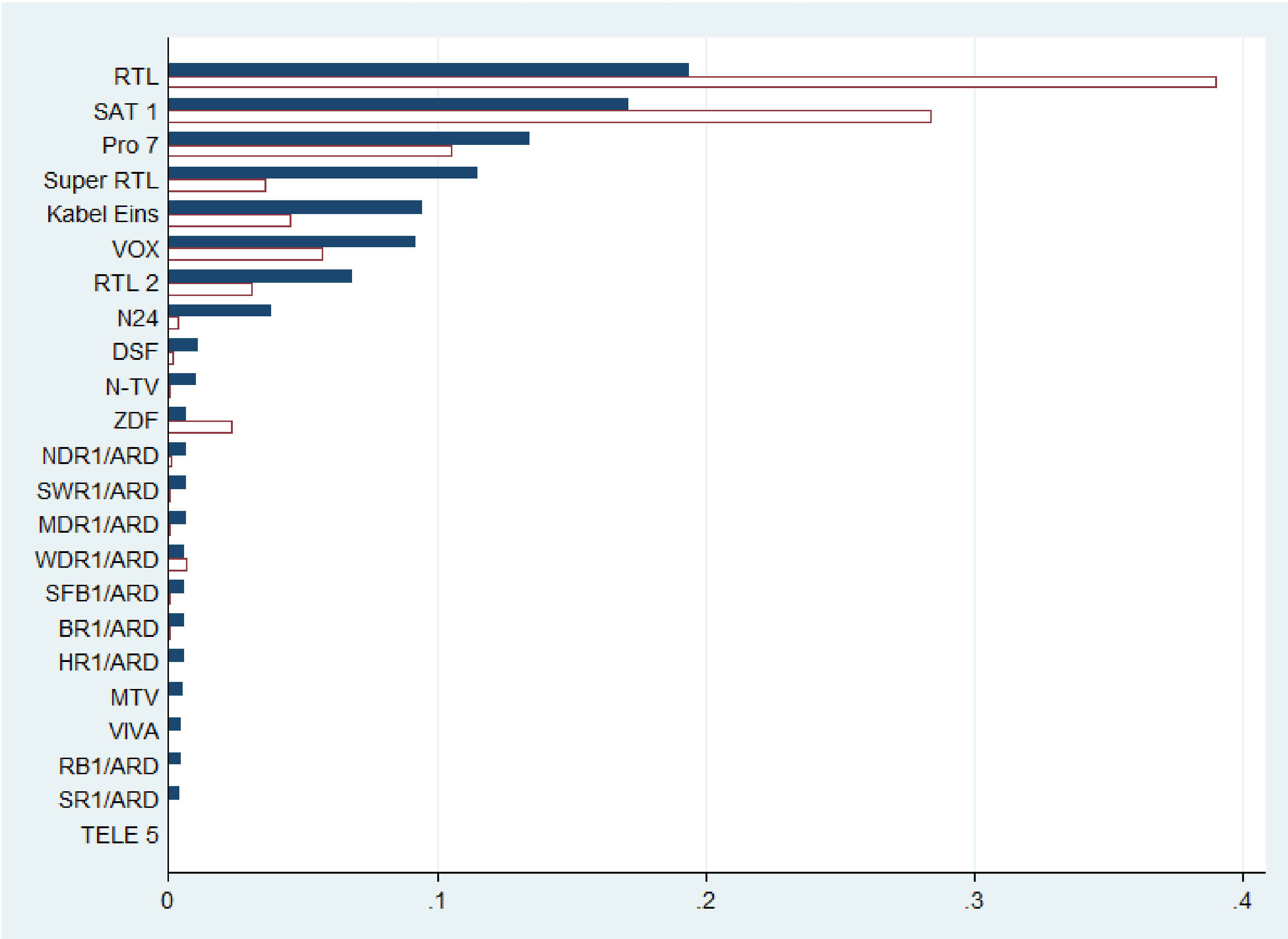
Turning attention to the advertising behavior, we need to disentangle the season effect on chocolate sales from the advertising effect. Figure (4) presents the number of ads and ad views each week for 2005. While the advertising levels are similar around Easter and Christmas, the purchase behavior in figure (2) peaks around Christmas. We can also see that the number of views matches approximately with the number of ads, although the ads over the summer months are viewed less frequently than the ads over the holidays. This suggests heterogeneity in both purchase behavior and advertising exposure across weeks.

Figure 4: Fraction of Annual Ads (solid bars) and Views (hollow bars) of Chocolate by Week, 2005



The data also show that chocolate advertising targets specific households. Figure (5) lists the ads per television station. The top four stations advertising chocolate in Germany are RTL, SAT 1, Pro 7, and Super RTL, Ads on RTL and SAT 1 are viewed more frequently than other channels. Super RTL is a kids channel, which explains the high number of ads broadcast on it. Certain households, such as those with young kids, might have a proclivity to purchase more chocolate as well as be exposed to more ads.

Figure 5: Fraction of Annual Ads (solid bars) and Views (hollow bars) of Chocolate by Channel, 2005



We find further evidence of ad targeting looking at the time slots chocolate advertisers run their ads. Figure (6) compiles all the chocolate advertising in 2005 and the majority of ads occur in the afternoon during the workweek. There is also a heavy concentration of ads during the week in the evenings from 5pm to 11pm, and Saturday throughout the day. Figure (7) counts the number of views, the majority of which occur during the week after 7pm. Ad exposure is clearly different than the overall advertising levels, and will likely be correlated with a household's characteristics and ad responsiveness.

Figure 6: Total Number of Chocolate Ads by Hour by Day, Summed across Weeks 2005

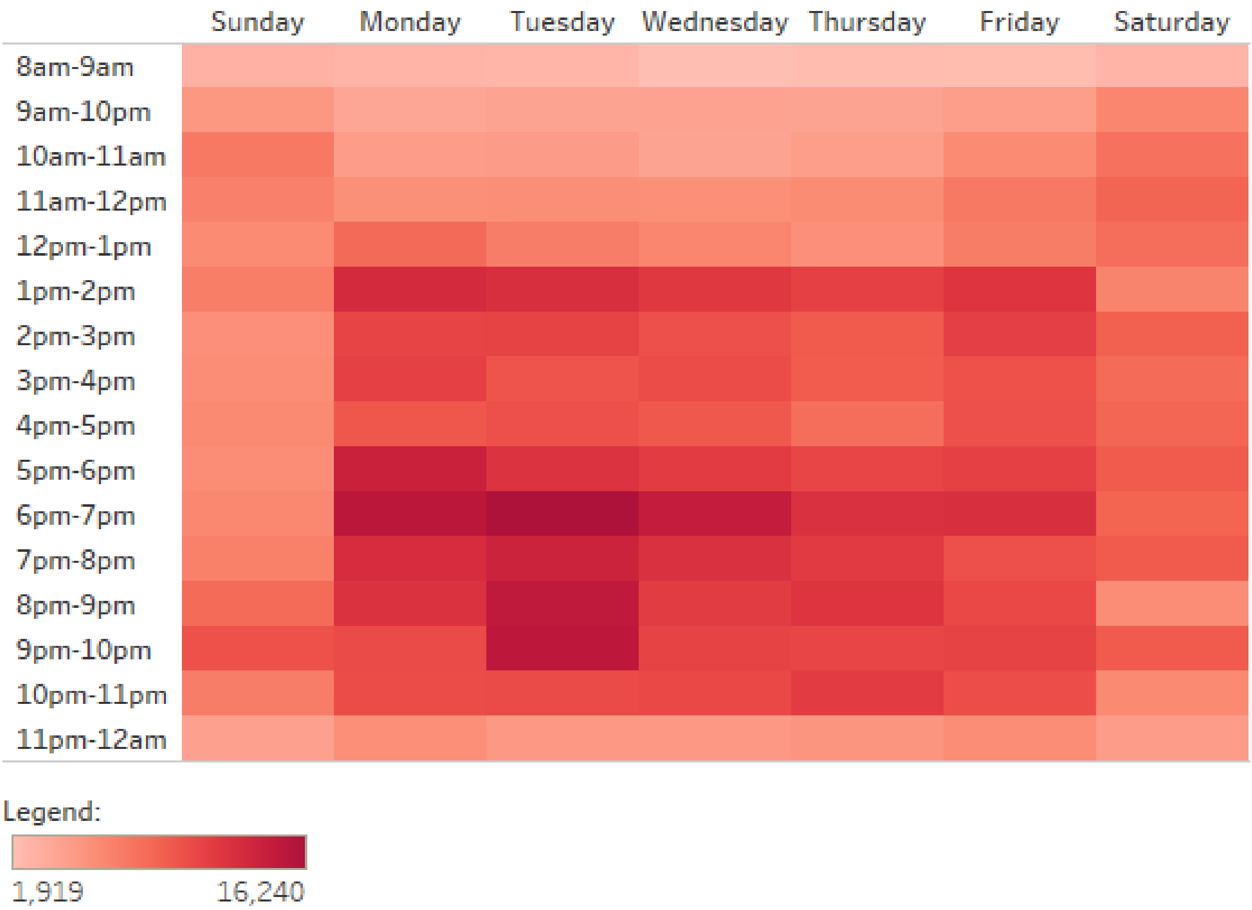
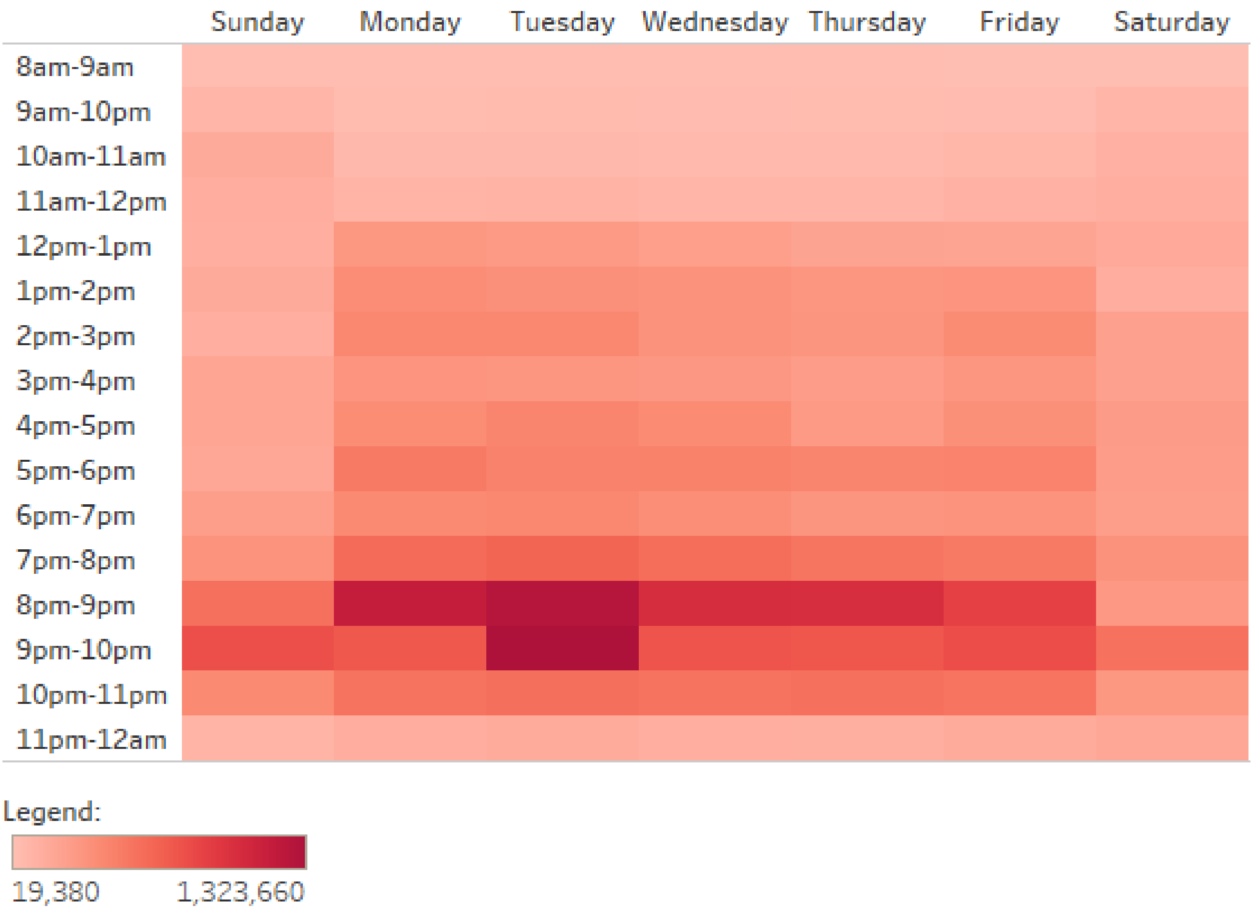


Figure 7: Total Number of Chocolate Views by Hour by Day, Summed across Weeks 2005



3.3 German Detergent Industry

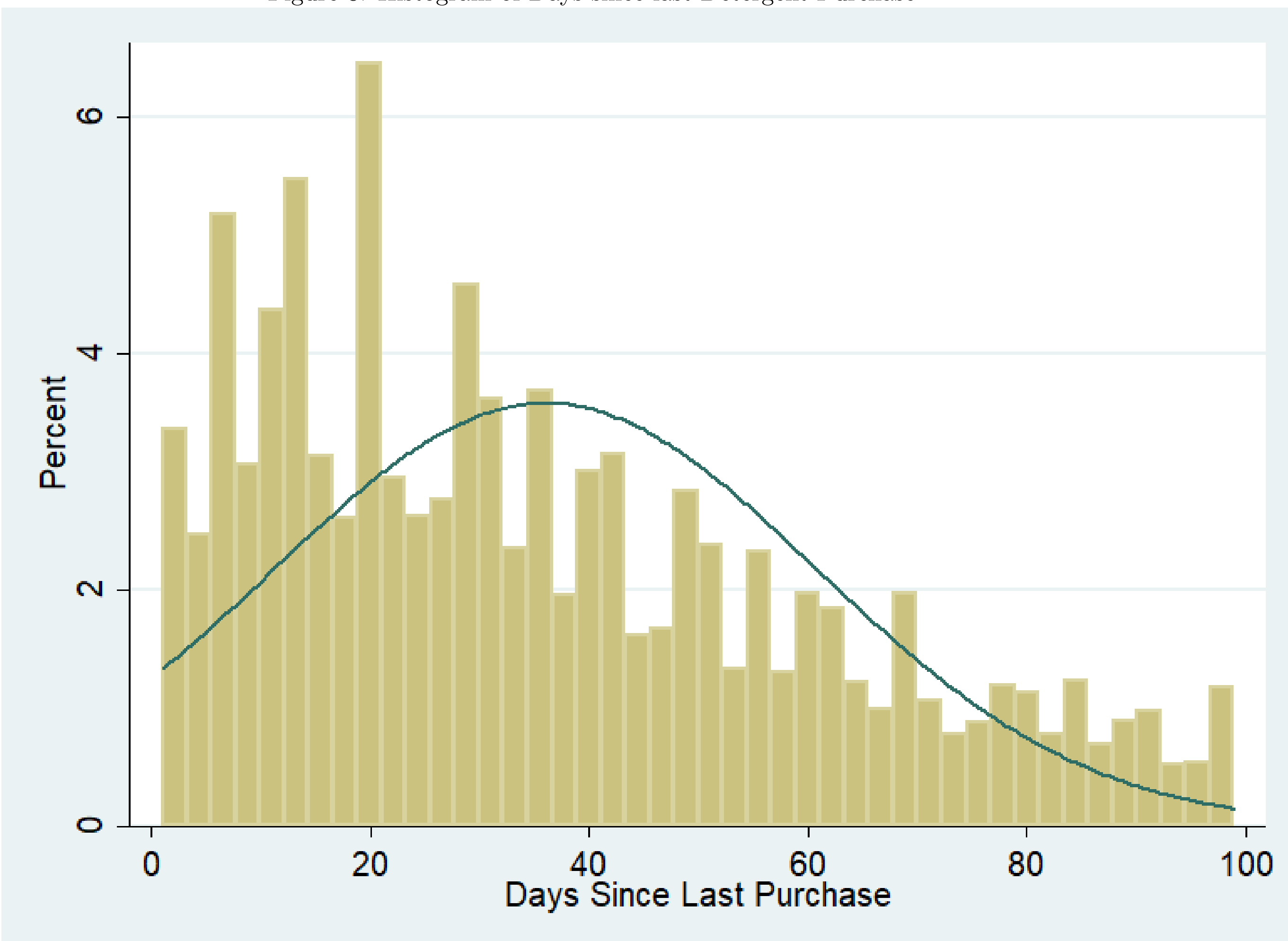
Over 90% of German households own a washing machine, and annual general laundry care sales have surpassed $1.75 billion each year since 2004. Over 75% of those sales are attributed to laundry detergent, with fabric softeners and laundry aids comprising the remaining 25%. Germans purchase a relatively high amount of powder laundry detergent, with the ratio of liquid to powder detergent close to 1:1 over the period of analysis.

Laundry detergent purchases occur twice as frequently at discount stores, compared to hypermarkets and supermarkets. This is reflected in the large share of private-label purchases. In our data, 47% of detergent purchase occasions result in the purchase of private label detergent. The large share of private labels is suggestive evidence that household heterogeneity is important to model. Because private labels are rarely advertised on television, a segment of the population may be relatively insensitive to television ads.

Household heterogeneity is a key part of our model.

The data, similar to chocolate, in an approximate sense come in continuous time: the day of purchase is recorded and the exact time the ad is viewed is recorded. However, laundry detergent is an infrequently purchased good; the time between purchases averages 36 days. Figure (8) shows that only 28% of households purchase detergent within 2 weeks, which is equivalent to the fraction of households which wait 11 weeks or more.

Figure 8: Histogram of Days since last Detergent Purchase



Unlike chocolate, detergent purchase behavior is not clustered around holidays although different seasons might require more or less laundry. Figure (9) shows that detergent purchases do not follow any specific pattern, although there are more purchases made in the end of the year.

Figure 9: Fraction of Annual Detergent Purchases by Week, 2005

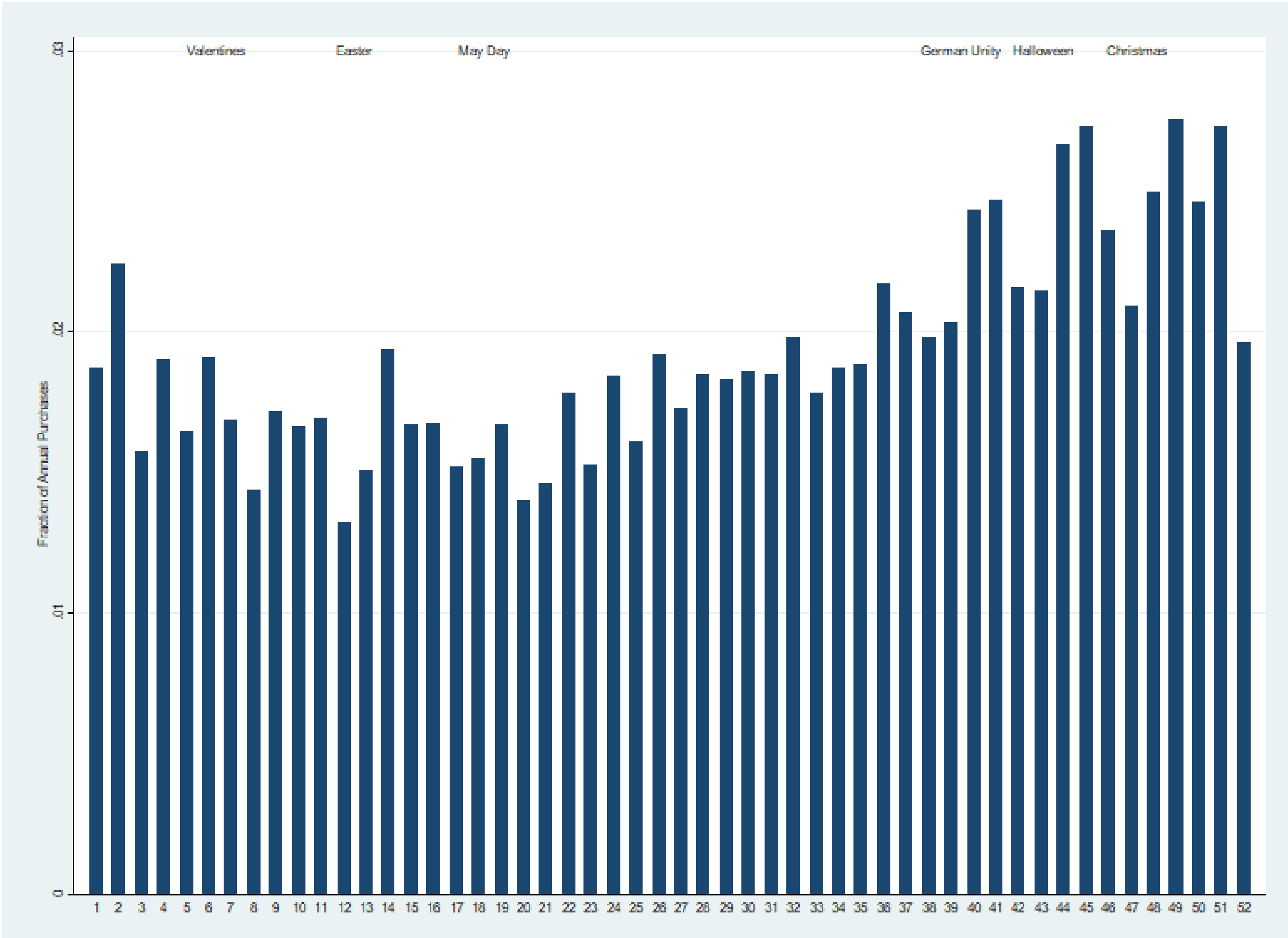


Figure (10) shows that monthly spending per purchase remains between 2-4 euros each month. Detergent is more expensive than chocolate, and it is purchased less frequently. It has a longer shelf-life and is not considered an "impulse purchase" to be quickly consumed. Interestingly, the advertising behavior of detergent purchases from figure (11) show that brands have distinct seasonal patterns about when they choose to advertise on television. Most brands do not advertise much in June and July, but throughout the rest of summer they advertise more than they do for the major holidays. This is very different from the purchase behavior, suggesting that these ads may not be driving purchase behavior despite the ad expenditure.

Figure 10: Boxplot of Average Monthly Detergent Spending per Purchase by Month, 2005

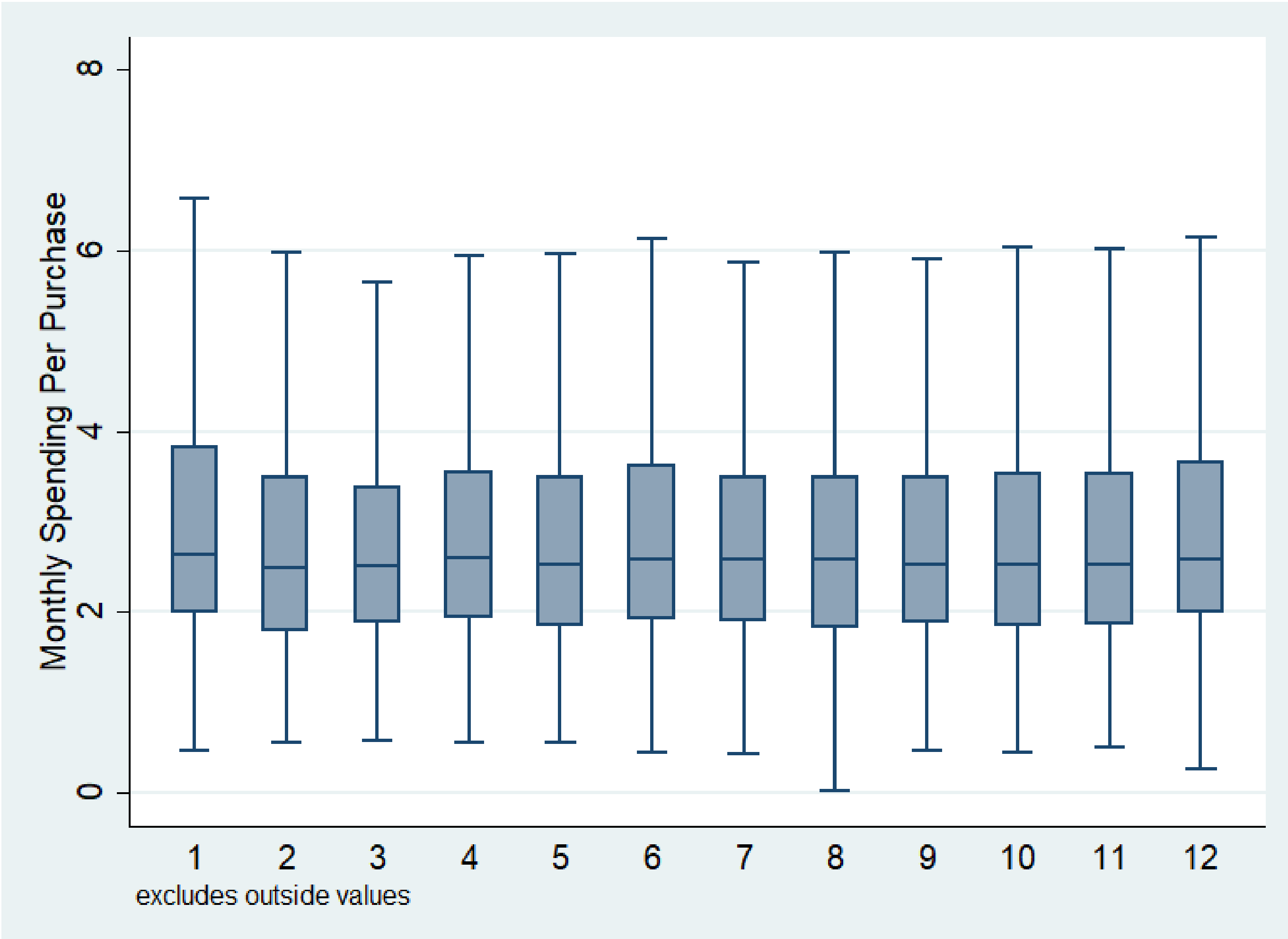
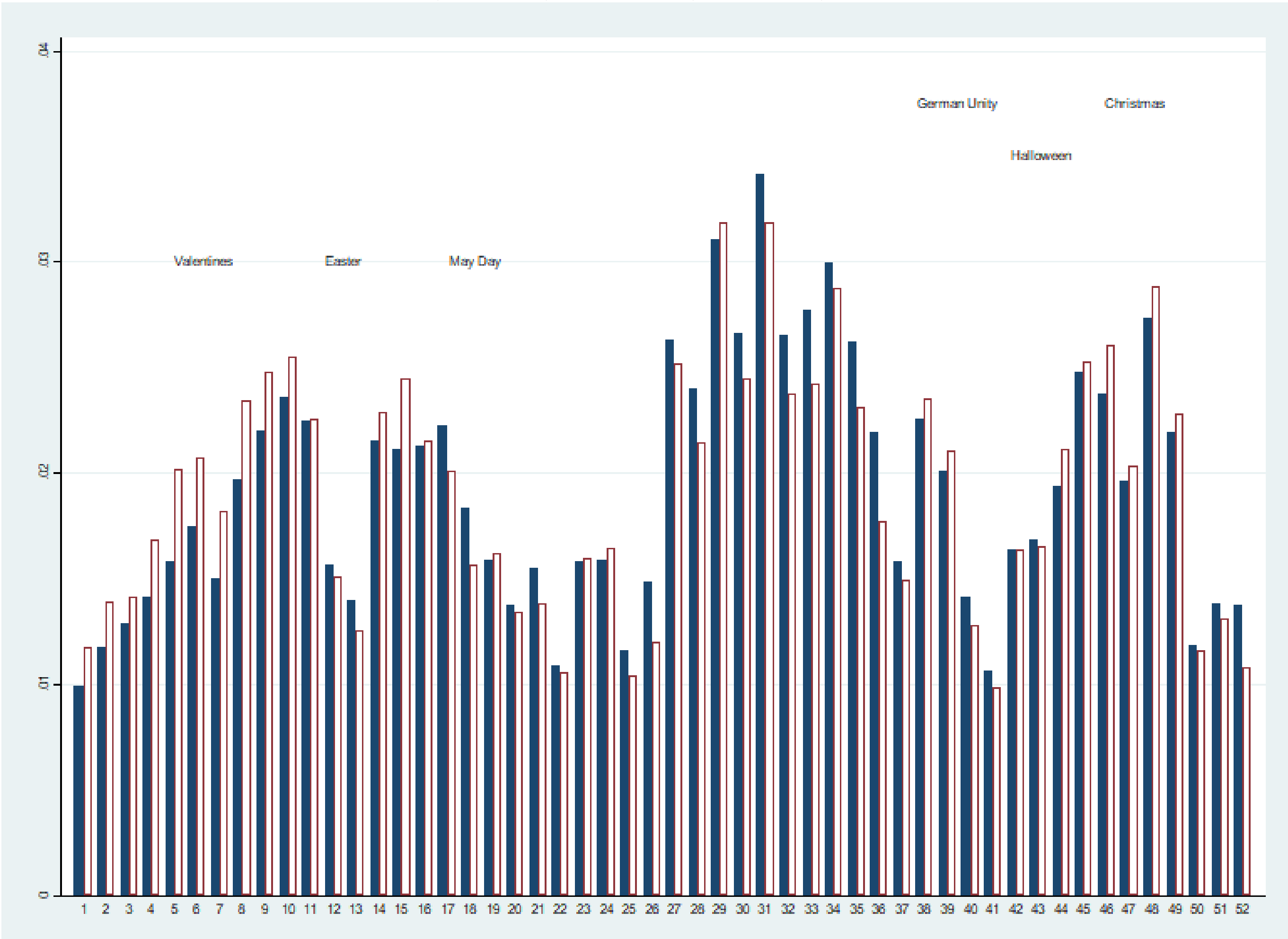


Figure 11: Fraction of Annual Ads (solid bars) and Views (hollow bars) of Detergent by Week, 2005



Detergent advertisers run their ads on less popular networks, which is markedly distinct from chocolate advertisers. Figure (12) has RTL and SAT1, similar to chocolate, but they also advertise heavily on VOX and Kabel Eins. They also go for a different demographic than chocolate ads, as the public channels account for over 5% of detergent's total viewership.

Figure 12: Fraction of Annual Ads (solid bars) and Views (hollow bars) of Chocolate by Channel, 2005

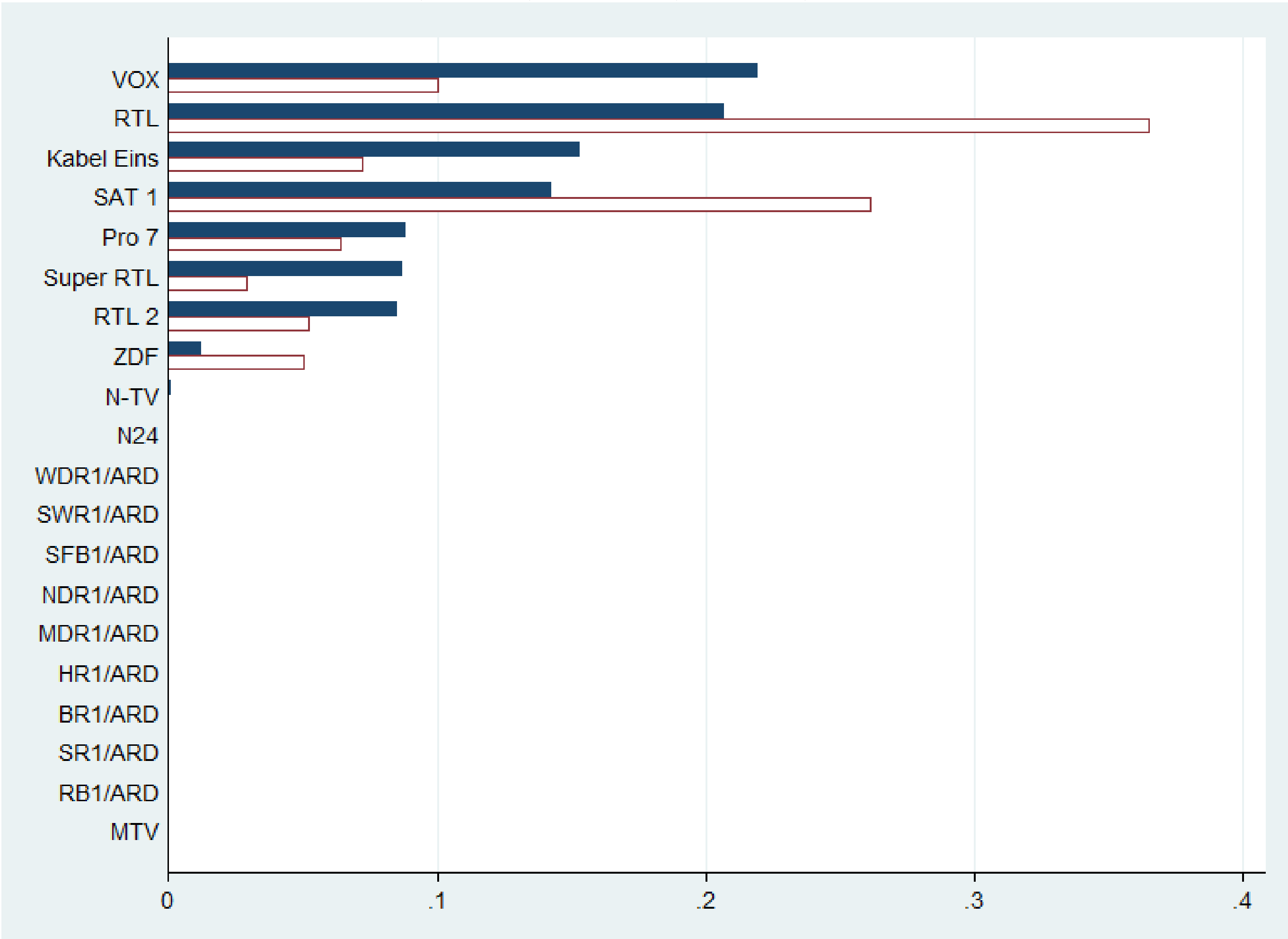


Figure (13) shows that detergent ads are heavily run from 9am to 5pm, which is different from chocolate advertising. The key demographic for laundry detergent might be housewives who watch TV during the day. Figure (14) shows that most of the views occur during the evenings, although there remains a significant viewership during the day as well.

Figure 13: Total Number of Detergent Ads by Hour by Day, Summed across Weeks 2005

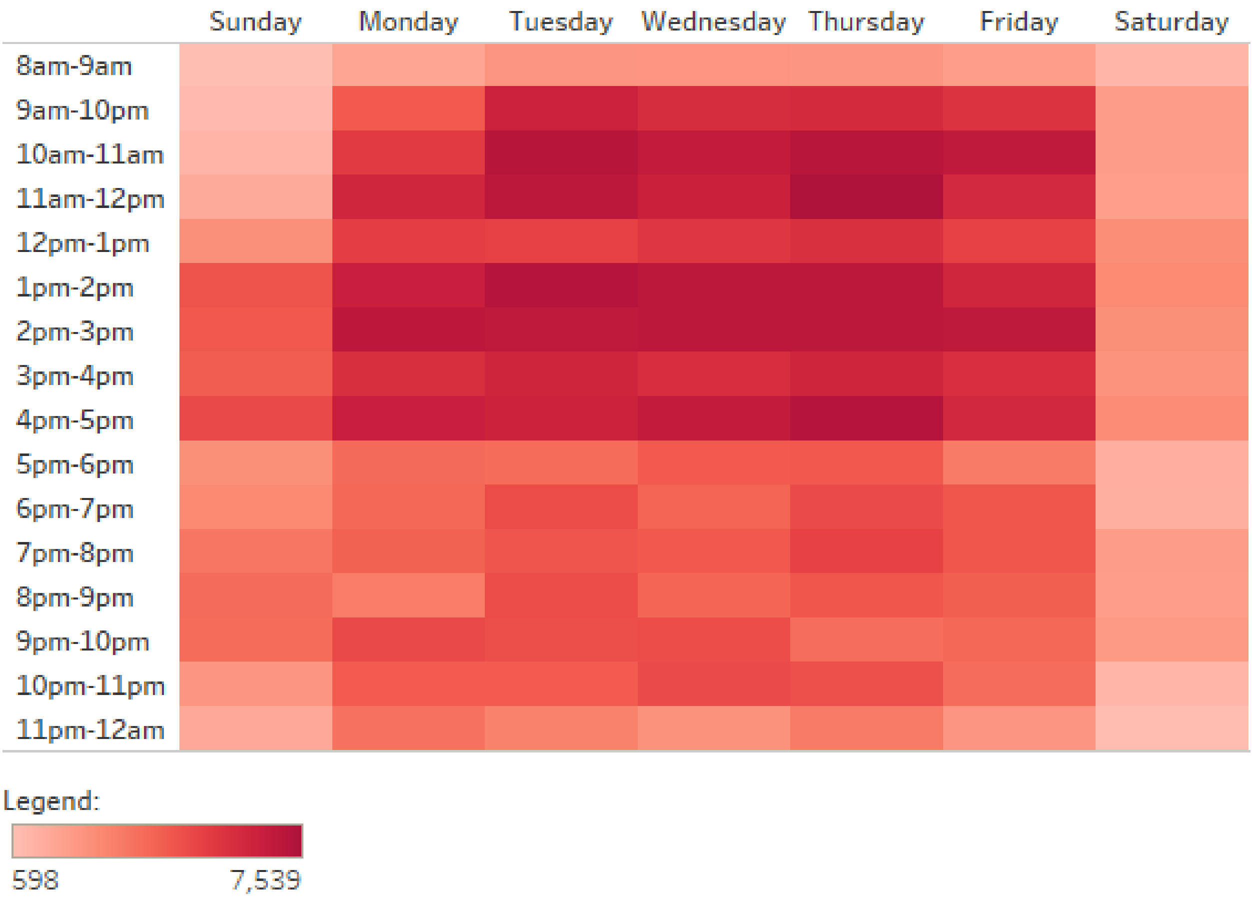
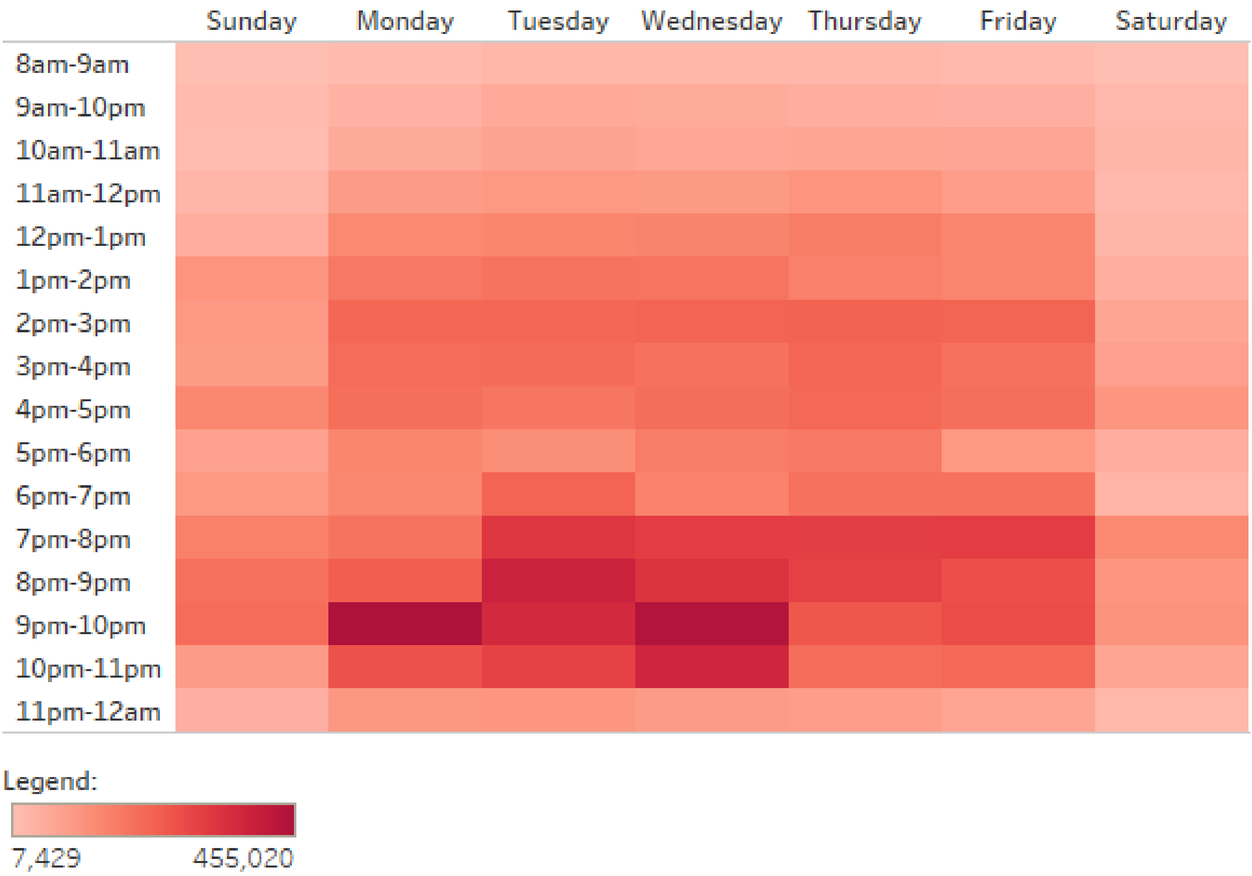


Figure 14: Total Number of Detergent Views by Hour by Day, Summed across Weeks 2005



3.4 German Stores

In the data we see 4 types of stores selling detergent and chocolate. 1) Supermarkets, which are typical grocery stores which specialize in both food and household products. 2) Discount stores, such as Aldi and Lidl, which operate similar to supermarkets and sell their goods often significantly discounted. They can accomplish this by avoiding expensive brand name goods, and often sell a much higher percentage of private label goods. 3) Hypermarkets, such as Walmart or Carrefour, have a wider range of merchandise than supermarkets, often acting as a "one-stop-shop." They sell high-volume and low-margin and often carry hundreds of thousands of brands. 4) Drug Stores have the most narrow range of merchandise available, and often carry a limited number of brands.

The major stores, the type of store, and the percentage of overall detergent and chocolate sales are listed in table (2). For data confidentiality, we do not disclose the store name, but it is clear that discount stores make up the largest sellers of both detergent and chocolate. Furthermore, discounters, hypermarkets, and supermarkets have relatively similar sales across the two industries, but drug stores sell much more detergent than chocolate.

Table 2: Market Share and Private Label Percentage for Chocolate and Detergent by Store

Store Name	Type of Store	% of Detergent Purchases	Private-Label % Detergent	% of Chocolate Purchases	Private-Label % Chocolate
Store 1	Discounter	12.82%	100.00%	13.27%	91.51%
Store 2	Discounter	10.32%	75.14%	10.85%	56.06%
Store 3	Discounter	8.27%	100.00%	11.60%	91.41%
Store 4	Discounter	2.32%	37.44%	3.70%	24.80%
Store 5	Discounter	3.24%	83.45%	5.08%	37.92%
Store 6	Supermarket	2.72%	41.65%	4.69%	48.10%
Store 7	Supermarket	1.22%	23.00%	2.86%	9.98%
Store 8	Hypermarket	7.85%	4.07%	7.81%	9.05%
Store 9	Hypermarket	5.35%	14.04%	4.95%	4.24%
Store 10	Hypermarket	0.80%	18.92%	1.33%	0.70%
Store 11	Drug Store	9.83%	27.36%	0.95%	0.74%
Store 12	Drug Store	6.14%	56.70%	0.11%	0.92%
Other		29.10%	19.23%	32.80%	5.93%
Industry Average			46.83%		37.09%

The data shows that discount stores control a large share of both the detergent and the chocolate sales. Discount stores sell a much higher fraction of private label goods while hypermarkets sell the lowest fraction of private label. There also appears to be more private label sales of detergent than chocolate, with 47% of detergent sales are private label compared to 37% of chocolate sales. There is much variation between store types as some discounters sell mostly brand-name goods, while some supermarkets may sell a higher fraction of private label. Furthermore, high private label sales in one industry does not necessarily mean that a store will sell a high fraction of private label in another industry. We can also see that in general the percentage of total sales for chocolate and detergent is relatively steady for the same store with the exception of drug stores, which sell a high share of detergent and a low share of chocolate.

3.5 Private-Label Goods

Private-label goods, also known as store brands, have the most success in commodity-driven, high purchase categories. In general, they compete in categories in which consumers do not perceive high levels of product differentiation, and they tend to grow at the expense of small- to mid-size brands. The categories in which private-label goods have succeeded vary country to country, but private labels have been consistently increasing their market share world wide. This is a dramatic shift, as private-label goods once carried the stigma of “cheap” and “low quality,” they are now seen as reasonable alternatives to brand-name goods.

A large driver for consumers is price, as stores are able to observe all the prices of brand-name goods and price their private-label goods accordingly. However, quality is a growing concern, and the most successful private-labels compete with (or surpass) the quality of brand name goods. A grocery store may even introduce multiple private label brands in the same category to compete for consumers across multiple segments. This trend is not restricted to just grocery stores–Amazon sells over 70 private-label brands across furniture, clothing, electronics, and food. Building trust with consumers through private-label brands can help attract loyal customers who may purchase goods in other product categories as well.

Beyond price, there are several characteristics of successful private-label product categories. A large determinant is the frequency which consumers purchase a product. With frequent purchases, consumers tend to be more price sensitive and notice changes in price. Furthermore, consumers might be willing to experiment with new brands given the "low-risk" of a fast purchase cycle.

Another key determinant is a low innovation rate for the product. If brand names are investing heavily in research and design of new products, private-label goods are less likely to compete. Commodity goods, such as milk and bread, represent low innovation goods that rarely change year to year. New products often represent less than 0.5% of yearly sales, which allow private-label goods to compete without a large R&D investment.

If products within a category are not highly differentiated, then it allows for private label to compete without introducing a gambit of products. For example, shampoo has not been successful as a private-label, and consumers tend to seek out a specific brand and category to address their needs. Along with differentiated products, high brand equity and marketing tend to discourage private-label sales in certain products. Industries which advertise (shampoo, toothpaste, etc) tend to sell higher levels of brand name goods than industries which do not heavily advertise (milk, eggs, bread).

Private-label goods compete across nearly all categories. While the characteristics of generally successful categories define most markets, private labels are clearly not restricted to those. This paper focuses on the puzzle that shows private labels can successfully compete against brand-names which advertise heavily. Chocolate and detergent are two product categories which advertise, have strong brand equity, innovate, and differentiate, yet private label goods control over one third of the market.

4 Household Demographics and Data

Our study is motivated by access to data from Nielsen Germany matching information on television ad exposure with information on laundry detergent and chocolate purchases. We focus on the nearly 4,000 households in the matched data who buy chocolate or detergent at least once. The data cover July 2004 through June 2006, a period of time where internet ads and internet purchases played less of a role in Germany than they do today.

A monitoring device is installed on the television in the household. Nielsen can see what channel the television is tuned to and hence match that to records on the ads shown. Except for the internet studies mentioned previously, this is one of the few datasets used for academic research with individual or household ad exposure linked to purchase data.

As researchers, we have access to only the information on the ads, not the original television program being viewed or on ads for other product categories. These data prevent us from estimating a structural model of television program choice and hence incidental ad exposure. Also, the television monitoring device cannot detect whether a household member is in the room viewing the television (instead of using the bathroom, etc) and does not record the identity of the household member viewing the ad. This might be an issue if one household member makes purchase decisions and another household member views an ad. Our econometric method does not adjust for such measurement issues and so our estimates must be interpreted to include the effects mentioned above. We define ad exposure to a particular brand as the total time (in seconds) exposed to that brand.

The purchase information is an example of "homescan" data. After visiting a grocery or similar store, each household member uses a device to scan the product bar codes of the purchased goods. We observe many things about the purchase, including brand pack size and the total spending on each product. Many of the television ads focus on the brand. Therefore, our dependent variable of interest in this advertising study is the decision to purchase a certain consumption quantity of our chosen focal brands. For chocolate we measure the number of grams, and for detergent we measure the number loads. Ads in the detergent category do sometimes distinguish between liquid and powder detergents, although we prefer to focus on brand and not the finer distinction between powder and liquids. Each brand generally sells both liquid and powder detergents.

The data contain no information on price except as computed as the total spending on a product on a purchase occasion divided by the data's measure of the quantity purchased. Importantly, we cannot observe the prices in the store visited for detergent products not purchased. The data cover households geographically dispersed throughout Germany, so we rarely observe multiple households shopping at the same store. Because we lack information on price for goods that are not purchased, we do not include price in our semi-structural model of choice.

The data break purchases down into multiple named brands as well as private label purchases and (separately) a collective category for minor brands that are not private labels. Not all brands advertise on television. We end up focusing our ad measures on six large brands: two chocolate and two detergent brands which routinely advertise on television, and a private label of each product. We do not explicitly name the focal brands for confidentiality purposes, but they are the largest brands in terms of detergent and chocolate sales in Germany. While the majority of our brands are owned by a few global conglomerates, the strategic interactions among firms owning the major brands plays no particular role in our study, which focuses on estimating the advertising elasticity.

5 Model

5.1 Dynamic Dorfman-Steiner

The traditional Dorfman-Steiner theorem states that the optimal level of advertising expenditure for a static, profit-maximizing monopolist is found where the ratio of advertising to sales equals the ratio of advertising elasticity of demand and price elasticity of demand. We show that if lagged advertising and lagged purchases enter the consumer's demand function it will call for higher levels of ad expenditure.

Consider a model where the state variables are the lagged demand, q_ℓ, and the lagged spending on advertising, A_ℓ. The Bellman equation $H(q_\ell, A_\ell)$ is

$$H(q_\ell, A_\ell) = \max_{A,p}(p-c)q(p, A, A_\ell, q_\ell) - A + \delta\big(H(q(p, A, A_\ell, q_\ell), A)\big), \tag{1}$$

where p is the price of each good and c is the cost to produce it. Let δ denote the discount factor. The first order condition with respect to A,

$$(p-c)\frac{\partial q(p, A, A_\ell, q_\ell)}{\partial A} - 1 + \delta\left(\frac{\partial H(q(p, A, A_\ell, q_\ell), A)}{\partial q(\cdot)}\frac{\partial q(p, A, A_\ell, q_\ell)}{\partial A} + \frac{\partial H(q(p, A, A_\ell, q_\ell), A)}{\partial A}\right) = 0 \tag{2}$$

The first order condition with respect to p:

$$(p-c)\frac{\partial q(p, A, A_\ell, q_\ell)}{\partial p} + q(p, A, A_\ell, q_\ell) + \delta\frac{\partial H(q(p, A, A_\ell, q_\ell))}{\partial q(\cdot)}\frac{\partial q(p, A, A_\ell, q_\ell)}{\partial p} = 0 \tag{3}$$

The envelope conditions are:

$$\frac{\partial H(q_\ell, A_\ell)}{\partial q_\ell} = (p-c)\frac{\partial q(p, A, A_\ell, q_\ell)}{\partial q_\ell} \tag{4}$$

$$\frac{\partial H(q_\ell, A_\ell)}{\partial A_\ell} = (p-c)\frac{\partial q(p, A, A_\ell, q_\ell)}{\partial A_\ell} \tag{5}$$

From equation (3) we can derive the following:

$$(p-c)\frac{\partial q(\cdot)}{\partial p}+\delta\left[\frac{\partial H(q(\cdot),A)}{\partial q(\cdot)}\right]\frac{\partial q(\cdot)}{\partial p}=-q(\cdot)$$

We replace the term in brackets using the envelope condition in equation (4),

$$(p-c)\frac{\partial q(\cdot)}{\partial p}+\delta\left[(p-c)\frac{\partial q(\cdot)}{\partial q_\ell}\right]\frac{\partial q(\cdot)}{\partial p}=-q(\cdot),$$

and finally we join terms and divide both sides by p to get:

$$\frac{(p-c)}{p}\left(1+\delta\frac{\partial q(\cdot)}{\partial q_\ell}\right)=-\frac{q(\cdot)}{p}\frac{1}{\partial q(\cdot)/\partial p}=-\frac{1}{\varepsilon_p}.$$

Equation (2) gives us:

$$\frac{A}{pq(\cdot)}=\frac{A}{pq(\cdot)}\left((p-c)\frac{\partial q(\cdot)}{\partial A}+\delta\left(\left[\frac{\partial H(q(\cdot),A)}{\partial q(\cdot)}\right]\frac{\partial q(\cdot)}{\partial A}+\left[\frac{\partial H(q(\cdot),A)}{\partial A}\right]\right)\right) \tag{6}$$

If we plug in our envelope condition from equation (4) and (5), we get:

$$\frac{A}{pq(\cdot)}=\frac{A}{pq(\cdot)}\left((p-c)\frac{\partial q(\cdot)}{\partial A}+\delta\left(\left[(p-c)\frac{\partial q(\cdot)}{\partial q_\ell}\right]\frac{\partial q(\cdot)}{\partial A}+\left[(p-c)\frac{\partial q(\cdot)}{\partial A_\ell}\right]\right)\right) \tag{7}$$

We can distribute $\frac{A}{q(\cdot)}$ across the terms, noting that the advertising elasticity of demand is $\varepsilon_A=\frac{A}{q(\cdot)}\frac{\partial q(\cdot)}{\partial A}$ and the lagged advertising elasticity of demand is $\varepsilon_{A_\ell}=\frac{A_\ell}{q(\cdot)}\frac{\partial q(\cdot)}{\partial A_\ell}$. Furthermore, in an equilibrium, $A=A_\ell$. We then have:

$$\frac{A}{pq(\cdot)}=\frac{(p-c)}{p}\left(\varepsilon_A+\delta\left(\left[\frac{\partial q(\cdot)}{\partial q_\ell}\right]\varepsilon_A+\varepsilon_{A_\ell}\right)\right) \tag{8}$$

and pulling ε_A out of the parenthesis,

$$\frac{A}{pq(\cdot)}=\frac{(p-c)}{p}\varepsilon_A\left(1+\delta\left(\left[\frac{\partial q(\cdot)}{\partial q_\ell}\right]+\frac{\varepsilon_{A_\ell}}{\varepsilon_A}\right)\right) \tag{9}$$

And re-arranging we have:

$$\varepsilon_A=\frac{A}{pq(\cdot)}\frac{p}{(p-c)}\frac{1}{\left(1+\delta\left(\left[\frac{\partial q(\cdot)}{\partial q_\ell}\right]+\frac{\varepsilon_{A_\ell}}{\varepsilon_A}\right)\right)} \tag{10}$$

Finally, we see that the ratio of our elasticities yields the familiar Dorfman-Steiner equation:

$$\frac{\varepsilon_A}{-\varepsilon_p} = \frac{A}{pq(\cdot)} \frac{1 + \delta \frac{\partial q(\cdot)}{\partial q_\ell}}{\left(1 + \delta \left(\left[\frac{\partial q(\cdot)}{\partial q_\ell}\right] + \frac{\varepsilon_{A_\ell}}{\varepsilon_A}\right)\right)} \tag{11}$$

where we note that in equilibrium (and assuming a non-negative advertising elasticity of demand),

$$\frac{1 + \delta \frac{\partial q(\cdot)}{\partial q_\ell}}{\left(1 + \delta \left(\left[\frac{\partial q(\cdot)}{\partial q_\ell}\right] + \frac{\varepsilon_{A_\ell}}{\varepsilon_A}\right)\right)} \leq 1.$$

So given that quantity demanded depends on both the previous quantity demanded and previous advertising,

$$\frac{\varepsilon_A}{-\varepsilon_p} \leq \frac{A}{pq(\cdot)},$$

which means that we would expect to spend even more money on advertising because it has a dynamic effect beyond the current advertising elasticity, ε_A. We can see, however, that as the lagged advertising elasticity of demand ε_{A_ℓ} or our discount factor δ approaches zero we return to the static environment with

$$\frac{\varepsilon_A}{-\varepsilon_p} = \frac{A}{pq(\cdot)}.$$

In the appendix we provide derivations when the state variable is only lagged purchases or only lagged advertising. The results are encompassed by this joint example in that lagged advertising will increase the advertising expenditure, while lagged purchase decisions do not affect the optimal level of spending. It may seem counterintuitive that a model which only incorporates lagged purchase decision will does not call for more advertising expenditure. However, lagged purchase decisions scale the advertising elasticity of demand and the price elasticity of demand by equal amounts, so the ratio will not change.

5.2 Model to Estimate

To identify the advertising elasticity of demand and address the sources of endogeneity discussed in the introduction, we use a correlated random coefficients model. Suppressing notation for a particular brand b, our model is

$$y_{it} = \alpha^*_{0,it} + \alpha_1 y_{i,t-1} + \beta^*_{0,it} x_{it} + \beta^*_{1,it} x_{i,t-1} + \gamma_t + u_{it}, \tag{12}$$

where y_{it} denotes household i's purchase decision quantity of a particular focal brand b in period t and x_{it} denotes household i's advertising exposure to brand b in period t. The period fixed effect, γ_t, denotes a brand's time varying intercept while u_{it} is an independent error. The three random coefficients, $\alpha^*_{0,it}$, $\beta^*_{0,it}$,

$\beta^*_{1,it}$, are defined as

$$\begin{aligned}\alpha^*_0 =& \quad \alpha_0 + \nu_{it},\\ \beta^*_{0,it} =& \quad \beta_0 + \xi_{it},\\ \beta^*_{1,it} =& \quad \beta_1 + \delta_{it}.\end{aligned}$$

Finally, the errors on the random coefficients, ν_{it}, ξ_{it} and δ_{it}, are all AR(1) and unobservable to the econometrician. Let them be denoted as:

$$\begin{aligned}\nu_{it} =& \quad \rho_1\nu_{i,t-1} + v_{it},\\ \xi_{it} =& \quad \rho_2\xi_{i,t-1} + e_{it},\\ \delta_{it} =& \quad \rho_3\delta_{i,t-1} + r_{it},\end{aligned}$$

where v_{it}, e_{it}, and r_{it} are innovations. We can re-express our equation as:

$$y_t = \alpha_0 + \alpha_1 y_{i,t-1} + \beta_0 x_{it} + \beta_1 x_{i,t-1} + \gamma_t + [\xi_{it}x_{it} + \delta_{it}x_{i,t-1} + \nu_{it} + u_{it}] \tag{13}$$

Where the term in brackets is unobservable to the econometrician. In what follows, I ignore the independent error, u_{it} because we cannot separately identify its moments from ν's innovations, but it can be added in without affecting the results of our other parameters. Additionally, the intercept α_0 cannot be separately identified from the time fixed effects, so we join those terms in estimation. All the parameters are unique to a particular brand, and we allow correlation between random coefficients across brands and products.

5.3 Isolating the Innovations

It is clear that equation (13) suffers from endogeneity as the measured variables appear in the unobserved error term. Furthermore, we cannot make use of Arellano-Bond type instruments as lagged x and y may be correlated with the AR(1) errors ν_{it}, ξ_{it} and δ_{it}, as those errors contain the entire history of innovations. In order to isolate the innovations and remove the AR(1) errors we must difference.

For exposition, assume we have a simple model:

$$y_t = \beta^*_0 x_{it} = \beta_0 x_{it} + [\xi_{it}x_{it}].$$

If we take a first difference, premultiplied by ρ_2, we have

$$y_t - \rho_2 y_{t-1} = \beta_0 x_{it} - \rho_2\beta_0 x_{i,t-1} + [\xi_{it}x_{it} - \rho_2\xi_{i,t-1}x_{i,t-1}],$$

where the term in brackets is the unobservable error. We then condition on the observations which have

$x_{it} = x_{i,t-1}$,

$$y_t - \rho_2 y_{t-1} = x_{it}\beta_0(1-\rho_2) + [x_{it}(\xi_{it} - \rho_2\xi_{i,t-1})].$$

Given that $\xi_{it} = \rho_2\xi_{i,t-1} + e_{it}$, we arrive at

$$y_t - \rho_2 y_{t-1} = x_{it}\beta_0(1-\rho_2) + [x_{it}e_{it}]. \tag{14}$$

While x_{it} is an endogenous variable, we can use lags of x_{it} and y_{it} from $t-2$ and back as instruments. We must perform a separate difference for each ρ parameter as well as impose conditioning restrictions for all the AR(1) variables which do not enter additively (in our model, only ν enters additively).

6 Estimation

To estimate the model, we first remove all the AR(1) errors by taking three differences. We begin with equation (12) and take a $d1$ period difference premultiplied by ρ_1^{d1} to remove the additive error, ν_{it}. Next we take a $d2$ period difference premultiplied by ρ_2^{d2} to remove the ξ_{it} term. Finally, we a a $d3$ period difference premultiplied by ρ_3^{d3} to remove the δ_{it} term. The choice of $d1$, $d2$, and $d3$ are important; the wrong differencing values, such as $d1 = d2 = d3$, will not lead to identification of the parameters. We generally achieve identification if no combination of the variables can sum to the same value as a different combination of the variables. For estimation, we choose $d1 = 2$, $d2 = 4$, and $d3 = 8$, which gives us the following conditional restrictions:

$$\begin{aligned} x_{t-4} &= x_t, & x_{t-9} &= x_{t-1}, \\ x_{t-6} &= x_{t-2}, & x_{t-11} &= x_{t-3}, \\ x_{t-12} &= x_{t-8}, & x_{t-13} &= x_{t-5}, \\ x_{t-14} &= x_{t-10}, & x_{t-15} &= x_{t-7}. \end{aligned}$$

We approximate these conditions using kernel weights in order to satisfy the requirements. Due to the generality of our model which allows the ρ variables to differ between the AR(1) errors, we require these 8 constraints and 15 periods of data to create one expression free from AR(1) errors. With fewer AR(1) errors or setting ρ to be equal for certain AR(1)'s, we could ease the number of periods needed and the conditioning requirements.

The differenced model still suffers from endogeneity as our measured variables (which include lags of y and current and lagged x's) may be correlated with the innovations. To address this we first need to identify valid instruments Z. Given that our differenced equation contains innovations going back to period $t-15$, we can use as instruments all lags of x and y from period $t-15$ and earlier. This follows the assumption that innovations may be correlated with current and future measured variables, but they are not correlated with measured variables from previous periods.

With the set of endogenous variables and instruments, we can apply Masten and Torgovitsky (2016) to identify the average partial effect of advertising on purchase decisions. In particular, we assume that the endogenous variables are exogenous conditional on a particular realization of some unobserved V. We assume that our endogenous variables are strictly increasing in V, such that $X = h_k(Z, V_k)$ where k denotes some quantile of V. We include the k notation to express that the effect of the instruments on X is not homogeneous and may change depending on different realizations of the unobserved V term.

We next need to solve for the rank statistic of each observation for each endogenous variable. We can use the knowledge that X is strictly increasing in V to invert that function and get the rank statistic given the set of endogenous variables and instruments. In practice we solve for the rank statistic by running a quantile regression of each endogenous variable on our instruments for a large range of quantiles. The rank statistic is then computed by finding where the endogenous variable is located compared to all the predicted values of the endogenous variable at each quantile. Once we solve for the rank statistic, we can condition on the realization of the rank statistic and given the assumptions of the model, the endogenous variables are no longer endogenous. For more details, see Masten and Torgovitsky (2016).

We are interested in the mean, variance, and covariances of our random coefficient variables. This requires the evaluation of three separate equations. The first is the simple linear model as in equation (14) which allows identification of the ρ's, β's, α, and the γ's. In order to identify the second moments of our innovations, we perform the same procedure, except now our equation of interest is squared. The interactions between the innovations are variances and covariances if they occur in the same period and zero otherwise. Then conditioning on the rank statistic, we are able to solve for the variances and covariances of the innovations for a particular brand. Finally, we need to solve for the covariances of the innovations between different brands. To do this, we multiply the two equations together and apply the same assumption as before: the covariance between innovations are zero if the innovations are draw in different periods.

The procedure above will supply us with the moments of the innovations and the point estimates of our parameters. We have not explicitly solved for the variance covariance matrix of the random coefficients yet, but we can use the moments of our innovations and the ρ parameters to get them. We treat the expression as a VAR system, and following the details in the appendix we can solve for the unconditional mean and variance of our random coefficients.

7 Results

There are two main results we are interested in: the advertising elasticity of demand and the covariance of ad elasticities across brands and products. The advertising elasticity of demand would need to be around .2 to justify the advertising spending in the detergent industry, and .1 to to justify the advertising spending in the chocolate industry. Below we can compare our result to “naive” approaches of 1) completely ignoring any potential endogeneity, and 2) controlling for potential endogeneity using an Arellano-Bond estimator, but ignoring the random coefficients.

The other result of interest is the correlation of the advertising elasticity of demand across brands and industries using the correlated random coefficients. This is a central feature of our technique which provides new insights helps drive firm strategy as we describe below.

7.1 Calculating the Advertising Elasticities of Demand

The advertising elasticity of demand is the percentage change in quantity demanded for a percent change in advertising. Given some demand function, $q(\cdot)$, and an advertising level A, we can calculate the advertising elasticity of demand as

$$\varepsilon_A = \frac{A}{q(A)}\frac{\partial q(A)}{A}.$$

Given parameter estimates of the demand function, we can get point estimates of the advertising elasticity of demand for the TV viewing population, as well as apply the delta method to get confidence intervals. We begin with the estimates for the naive model in table (3), which assumes no endogeneity issues.

Table 3: Naive Estimates of Ad Elasticity for Chocolate and Detergent Brands

	Coef.	Std. Err.	z	P>z	[95% Conf.	Interval]
Chocolate Brand 1:	0.05	0.02	2.62	0.01	0.01	0.09
Chocolate Brand 2:	0.02	0.03	0.72	0.47	-0.03	0.07
Detergent Brand 1:	0.03	0.09	0.32	0.75	-0.15	0.21
Detergent Brand 2:	0.21	0.09	2.28	0.02	0.03	0.38

The naive results show that the ad spending might be justified for certain brands. The 95% confidence interval shows that chocolate brand 1 and detergent brand 2 are both positive and significant, and the detergent brand 2 achieves the .2 ad elasticity needed. However, given the endogeneity concerns, these results are likely biased. Next, we take into account potential endogeneity issues using an Arellano-Bond estimator with lagged observations and purchases as instruments table (4).

The Arellano-Bond estimator performs GMM on a differenced equation to eliminate fixed effects which are inconsistent if strict exogeneity fails. Immediately, the point estimates fall and none of the confidence

Table 4: Arellano-Bond Estimates of Ad Elasticity for Chocolate and Detergent Brands

	Coef.	Std. Err.	z	P>z	[95% Conf.	Interval]
Chocolate Brand 1:	-0.04	0.01	-2.88	0.00	-0.06	-0.01
Chocolate Brand 2:	0.01	0.01	1.06	0.29	-0.01	0.04
Detergent Brand 1:	-0.17	0.01	-24.55	0.00	-0.19	-0.16
Detergent Brand 2:	0.36	0.00	379.54	0.00	0.35	0.36

intervals even approach our threshold levels of advertising elasticity needed. While Arellano-Bond addresses some potential endogeneity between time-persistent household characteristics and advertising, it does not address the full spectrum of endogeneity we are concerned with, such as time-specific household demand shocks potentially being correlated with time-specific household ad exposures. We address these other sources of endogeneity using the correlated random coefficients model to estimate the advertising elasticities for those four chosen focal brands in table (5).

Table 5: Correlated Random Coefficients Estimates of Ad Elasticity for Chocolate and Detergent Brands

	Coef.	Std. Err.	z	P>z	[95% Conf.	Interval]
Chocolate Brand 1:	0.29	2.69	0.11	0.91	-4.98	5.57
Chocolate Brand 2:	-0.12	3.8	-0.03	0.97	-7.57	7.33
Detergent Brand 1:	0.54	28.65	0.02	0.98	-55.63	56.71
Detergent Brand 2:	1.15	28.77	0.04	0.96	-55.24	57.54

While a naive approach leads to the conclusion that the advertising elasticity of demand for certain products justifies the spending in the data, the results are confounded by the endogeneity issues mentioned in the introduction. The naive elasticities are much higher than the other models' results, suggesting the endogeneity bias is a large concern. The results between the Arellano-Bond estimator and our random coefficient model both show that the advertising elasticity of demand is much lower than the .2 needed for detergent and .1 needed for chocolate. The standard errors of the advertising elasticity of demand are higher for the correlated random coefficients model as the coefficients vary by household and time. The results suggests that for a subset of households, their responsiveness to advertising justifies the ad-spending, although the average advertising elasticity of demand is not significantly different from zero. This result is in line with the data, showing that a large fraction of consumers are unresponsive to advertising as they purchase the private label good regardless of ad exposure.

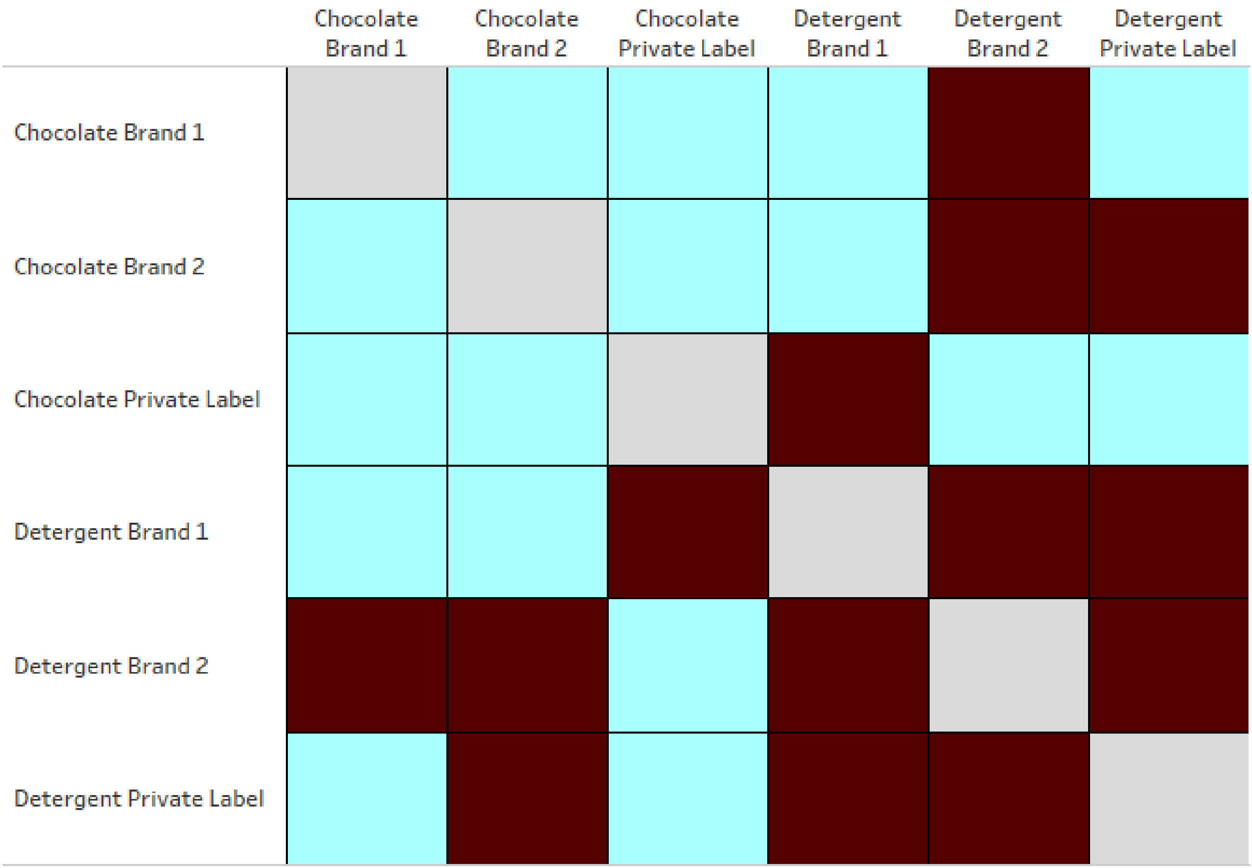
7.2 Correlation of Advertising Elasticity of Demand among Products and Brands

An important and powerful feature of our correlated random coefficient model is the ability to estimate correlation patterns between random coefficients of different products. Among other things, this allows us to estimate the correlation of the advertising elasticity of demand. This is important to firms for many reasons. First, a firm might want to know how households' advertising responsiveness varies across products in order to determine which products consumers view as substitutes. For example, if a product's ad elasticity was negatively correlated with the private label demand shock, then the firm would know that their segment of ad-responsive consumers is not being driven to private label as a substitute.

The correlated random coefficients model results in the correlation pattern of advertising elasticity shown in figure (15). We can glean a few important things from this. First, the correlation between the private label demand for chocolate is correlated with the private label demand for detergent, suggesting household stability of preferences. This tells us that consumers who purchase private label in one category are more likely to purchase it in another. Looking just at the chocolate industry, we see that when households are responsive to chocolate brand 1, they are also responsive to chocolate brand 2, as well as the private label. This implies that as a household's demand for private label chocolate increases, they are more responsive to advertising. They know they want chocolate and can be swayed by advertising to a particular brand.

The detergent industry has the opposite correlation pattern. Consumers which are responsive to a particular detergent brand are less likely to be responsive to the other brand. This suggests that an advertising campaign which shocks the ad elasticity may make consumers less responsive to the other brands as well as weaken the demand for private label.

Figure 15: Correlation Matrix for Ad Elasticities
Positive Correlation is Light, Negative Correlation is Dark



Some other firm strategies may involve strategic offense/defense. For example, if a brand is losing market share, the correlation between ad elasticities can show how to best address it. It identifies which brands' advertising consumers respond similarly to, and how the private label's demand fluctuates with household ad responsiveness.

8 Conclusion

This paper introduced a correlated random coefficients model which allows for flexible patterns of correlation across observed and unobserved variables and across different brands and products. This is a very powerful applied tool, as the correlated parameters can be used to direct firm strategy, both with respect to internal allocation of resources and to external threats and opportunities.

Our model extends the Arellano-Bond estimator by controlling for four very important sources of endogeneity: 1) Brand-Level advertising may be correlated with brand-specific, aggregate demand shocks. 2)

Household ad exposure may be correlated with household persistent tastes for a brand. 3) Time-specific household demand shocks may be correlated with time-specific household ad exposures. 4) Distinguishing between unobserved heterogeneity and state dependence. The results from our application show that both the correlated random coefficient model and Arellano-Bond estimator addressed the endogeneity that plagued a naive estimation of advertising elasticity of demand. In addition to controlling for richer patterns of endogeneity, our model allows for estimation of the correlation matrix of parameters, which contains very important information for firms.

The correlated random coefficient model can be used with much more precision depending on the application. For example it can handle the scenarios when 1) a firm sells multiple products in the same market, or 2) has multiple advertising campaigns, or 3) wants to consider every private label individually. The correlation matrix of the unobserved time-specific heterogeneity will contain correlations across every random coefficient across all possible brands. This indeed serves as a very useful tool and immediately reveals a rich pattern of underlying consumer behavior.

As a final thought, this model is not restricted to advertising and purchases. One potential application may be a firm that is interested in driving website traffic. If a firm can track individual customers across its websites and is interested in seeing what makes them engaged (number of seconds on site, or other outcome behavior), they can then measure the effect of various forms of treatment (text alerts, emailed coupons, etc) on the outcome and identify the correlation matrix between different websites. For example, consumers may respond similarly to sports and politics website engagement but those consumers do not respond similarly to cooking websites. A firm may use this to 1) increase engagement with a broader range of websites for the lowest cost. A firm can identify the treatments which have the highest correlated responsiveness among its websites and shift resources to these treatments. 2) Identify which competitor websites are true threats and which can be ignored. If a competitive website, with correlation in responsive among consumers, increases their engagement efforts they are not going to lure away many consumers. Finally, 3) identify potential websites to acquire in order to diversify. Once the correlation of the random coefficients between websites is determined, those sites with negative correlation primarily serve a different dimension of consumer who is not equally responsive to all brands, and an acquisition may yield a more diverse set of consumers. These are some examples of the strategic decisions that our method can advise on.

References

[1] Ackerberg, D. A. Empirically distinguishing informative and prestige effects of advertising. *RAND Journal of Economics* (2001), 316–333.

[2] Arellano, M., and Bond, S. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The review of economic studies 58*, 2 (1991), 277–297.

[3] Bagwell, K. The economic analysis of advertising. *Handbook of industrial organization 3* (2007), 1701–1844.

[4] Blake, T., Nosko, C., and Tadelis, S. Consumer heterogeneity and paid search effectiveness: A large-scale field experiment. *Econometrica 83*, 1 (2015), 155–174.

[5] Chernozhukov, V., and Hansen, C. An IV model of quantile treatment effects. *Econometrica 73*, 1 (2005), 245–261.

[6] Dubé, J.-P., and Manchanda, P. Differences in Dynamic Brand Competition Across Markets: An Empirical Analysis. *Marketing Science 24*, 1 (feb 2005), 81–95.

[7] Florens, J.-P., Heckman, J. J., Meghir, C., and Vytlacil, E. Identification of treatment effects using control functions in models with continuous, endogenous treatment and heterogeneous effects. *Econometrica 76*, 5 (2008), 1191–1206.

[8] Golden, J. M. Essays on Labor Economics and Advertising Auctions.

[9] Gordon, B. R., and Hartmann, W. R. Advertising competition in presidential elections. *Quantitative Marketing and Economics 14*, 1 (2016), 1–40.

[10] Hartmann, W. R., and Klapper, D. Super bowl ads. *Marketing Science 37*, 1 (2017), 78–96.

[11] Heckman, J., and Vytlacil, E. Instrumental variables methods for the correlated random coefficient model: Estimating the average rate of return to schooling when the return is correlated with schooling. *Journal of Human Resources* (1998), 974–987.

[12] Imbens, G. W., and Angrist, J. D. Identification and Estimation of Local Average Treatment Effects. *Econometrica 62*, 2 (1994), 467–475.

[13] Lewis, R. A., and Reiley, D. H. Online ads and offline sales: measuring the effect of retail advertising via a controlled experiment on Yahoo! *Quantitative Marketing and Economics 12*, 3 (2014), 235–266.

[14] Masten, M. A., and Torgovitsky, A. Identification of instrumental variable correlated random coefficients models. *Review of Economics and Statistics 98*, 5 (2016), 1001–1005.

[15] Nagel, C. Demand Models with Price Endogeneity and Advertising, 2010.

[16] Shapiro, B. T. Positive spillovers and free riding in advertising of prescription pharmaceuticals: The case of antidepressants. *Journal of Political Economy 126*, 1 (2018), 381–437.

[17] Stephens-Davidowitz, S., Varian, H., and Smith, M. D. Super returns to Super Bowl ads? *Quantitative Marketing and Economics 15*, 1 (2017), 1–28.

[18] Tellis, G. J. Advertising exposure, loyalty, and brand purchase: A two-stage model of choice. *Journal of marketing research* (1988), 134–144.

[19] Wooldridge, J. M. Further results on instrumental variables estimation of average treatment effects in the correlated random coefficient model. *Economics letters 79*, 2 (2003), 185–191.

A Appendix

A.1 Dorfman-Steiner Derivation in a Dynamic Environment with State Variable: Lagged Ad Expenditure

Given $H(A_\ell)$ where is A_ℓ the lagged spending on advertising, we arrive at:

$$H(A_\ell) = \max_{A,p} (p-c)q(p,A,A_\ell) - A + \delta\big(H(A)\big) \tag{15}$$

The first order condition with respect to A,

$$(p-c)\frac{\partial q(p,A,A_\ell)}{\partial A} - 1 + \delta\frac{\partial H(A)}{\partial A} = 0 \tag{16}$$

Where the term in red is the difference between the dynamic setting and the static one. The first order condition with respect to p:

$$(p-c)\frac{\partial q(p,A,A_\ell)}{\partial p} + q(p,A,A_\ell) = 0 \tag{17}$$

The envelope condition is:

$$\frac{\partial H(A_\ell)}{\partial A_\ell} = (p-c)\frac{\partial q(p,A,A_\ell)}{\partial A_\ell} \tag{18}$$

From equation (17) we can derive the following:

$$\frac{(p-c)}{p} = -\frac{q(p,A,A_\ell)}{p}\frac{1}{\partial q(p,A,A_\ell)/\partial p} = -\frac{1}{\varepsilon_p}$$

Multiplying equation (16) by $\frac{A}{pq(\cdot)}$ gives us:

$$\frac{A}{pq(\cdot)} = \frac{A(p-c)}{pq(\cdot)}\frac{\partial q(p,A,A_\ell)}{\partial A} + \delta\frac{A}{pq(\cdot)}\frac{\partial H(A)}{\partial A} \tag{19}$$

If we plug in our envelope condition from equation (18), we get:

$$\frac{A}{pq(\cdot)} = \frac{A(p-c)}{pq(\cdot)}\frac{\partial q(p,A,A_\ell)}{\partial A} + \delta\frac{A}{pq(\cdot)}(p-c)\frac{\partial q(p,A',A)}{\partial A} \tag{20}$$

In the dynamic world, once we have reached an equilibrium, $A_\ell = A = A'$. We can thus reexpress equation (20) as:

$$\frac{A}{pq(\cdot)} = \frac{(p-c)}{p}\varepsilon_A + \delta\frac{(p-c)}{p}\varepsilon_{A_\ell} \tag{21}$$

Where ε_A is the elasticity of demand with respect to advertising and ε_{A_ℓ} is the elasticity of demand with respect to lagged advertising. Finally we divide both sides by $\frac{(p-c)}{p}$ and we arrive at:

$$\varepsilon_A + \delta\varepsilon_{A_\ell} = \frac{A}{pq(\cdot)} \frac{p}{(p-c)} \tag{22}$$

Finally, we see that the ratio of our elasticities yields the familiar Dorfman-Steiner equation:

$$\frac{\varepsilon_A + \delta\varepsilon_{A_\ell}}{-\varepsilon_p} = \frac{A}{pq(\cdot)}$$

This tells us that the advertising to revenue ratio should actually be higher than the ratio of ad elasticity to price elasticity, on account of the effect of lagged advertising. We can also see that as $\delta \to 0$ we enter our static world.

A.2 Dorfman-Steiner Derivation in a Dynamic Environment with State Variable: Lagged Quantity Demanded

Given $H(q_\ell)$ where is q_ℓ is the lagged demand, we arrive at:

$$H(q_\ell) = \max_{A,p}(p-c)q(p,A,q_\ell) - A + \delta\big(q(p,A,q_\ell)\big) \tag{23}$$

The first order condition with respect to A,

$$(p-c)\frac{\partial q(p,A,q_\ell)}{\partial A} - 1 + \delta\frac{\partial H(q(p,A,q_\ell))}{\partial q(\cdot)}\frac{\partial q(p,A,q_\ell)}{\partial A} = 0 \tag{24}$$

Where the term in red is the difference between the dynamic setting and the static one. The first order condition with respect to p:

$$(p-c)\frac{\partial q(p,A,q_\ell)}{\partial p} + q(p,A,q_\ell) + \delta\frac{\partial H(q(p,A,q_\ell))}{\partial q(\cdot)}\frac{\partial q(p,A,q_\ell)}{\partial p} = 0 \tag{25}$$

The envelope condition is:

$$\frac{\partial H(q_\ell)}{\partial q_\ell} = (p-c)\frac{\partial q(p,A,q_\ell)}{\partial q_\ell} \tag{26}$$

From equation (25) we can derive the following:

$$(p-c)\frac{\partial q(\cdot)}{\partial p} + \delta\left[\frac{\partial H(q(\cdot))}{\partial q(\cdot)}\right]\frac{\partial q(\cdot)}{\partial p} = q(\cdot)$$

We replace the term in brackets using the envelope condition in equation (26):

$$(p-c)\frac{\partial q(\cdot)}{\partial p} + \delta\left[(p-c)\frac{\partial q(\cdot)}{\partial q_\ell}\right]\frac{\partial q(\cdot)}{\partial p} = q(\cdot)$$

and finally we join terms and divide both sides by p to get:

$$\frac{(p-c)}{p}\left(1+\delta\frac{\partial q(\cdot)}{\partial q_\ell}\right) = -\frac{q(\cdot)}{p}\frac{1}{\partial q(\cdot)/\partial p} = -\frac{1}{\varepsilon_p}$$

Equation (24) gives us:

$$\frac{A}{pq(\cdot)} = \frac{A}{pq(\cdot)}\left((p-c)\frac{\partial q(\cdot)}{\partial A} + \delta\left[\frac{\partial H(q(\cdot))}{\partial q(\cdot)}\right]\frac{\partial q(\cdot)}{\partial A}\right) \tag{27}$$

If we plug in our envelope condition from equation (26), we get:

$$\frac{A}{pq(\cdot)} = \frac{A}{pq(\cdot)}\left((p-c)\frac{\partial q(\cdot)}{\partial A} + \delta\left[(p-c)\frac{\partial q(\cdot)}{\partial q_\ell}\right]\frac{\partial q(\cdot)}{\partial A}\right) \tag{28}$$

Collecting terms we arrive at:

$$\frac{A}{pq(\cdot)} = \frac{p-c}{p}\varepsilon_A\left(1+\delta\frac{\partial q(\cdot)}{\partial q_\ell}\right) \tag{29}$$

$$\varepsilon_A = \frac{A}{pq(\cdot)}\frac{p}{p-c}\frac{1}{\left(1+\delta\frac{\partial q(\cdot)}{\partial q_\ell}\right)} \tag{30}$$

Finally, we see that the ratio of our elasticities yields the familiar Dorfman-Steiner equation:

$$\frac{\varepsilon_A}{-\varepsilon_p} = \frac{A}{pq(\cdot)}$$

A.3 Estimation of the Random Coefficient Covariance Matrix

Consider our set of the AR(1) term of the random coefficients, where I denote the brand by b,

$$\begin{aligned}\nu_{b,it} = &\quad \rho_{b,1}\nu_{b,i,t-1} + v_{b,it},\\ \xi_{b,it} = &\quad \rho_{b,2}\xi_{b,i,t-1} + e_{b,it},\\ \delta_{b,it} = &\quad \rho_{b,3}\delta_{b,i,t-1} + r_{b,it},\end{aligned}$$

Given that we have B brands, we have a set of $3 \times B$ equations, where we solved for all the ρ's and all the variance-covariance terms of the innovations v, e, and u. We can reexpress this system of equations as a VAR(1) process, which gives us

$$j_t = A_1 j_{t-1} + u_t$$

where j_t is a $3B \times 1$ vector corresponding to our $[\nu_{1,it}, \xi_{1,it}, \delta_{1,it}, \nu_{2,it}, \xi_{2,it}, \delta_{2,it}, \ldots, \nu_{B,it}, \xi_{B,it}, \delta_{B,it}]$, A_1 is an $B \times B$ diagonal matrix where the diagonals correspond to $[\rho_{1,1}, \rho_{1,2}, \rho_{1,3}, \ldots, \rho_{B,1}, \rho_{B,2}, \rho_{B,3}]$, j_{t-1} is the $3B \times 1$ vector of lagged variables, and u_t is the $3B \times 1$ vector of innovations corresponding to $[v_{1,it}, e_{1,it}, r_{1,it}, \ldots, v_{B,it}, e_{B,it}, r_{B,it}]$.

Our vector of innovations u_t satisfies the following:

1. $E[u_t] = 0$
2. $E[u_t u_t'] = \Sigma_u$, where Σ_u is the $3B \times 3B$ positive semidefinite contemporaneous covariance matrix
3. $E[u_t u_{t-k}] = 0$ for any non-zero k

We can then solve for the unconditional variance between the random coefficients. Given $j_t = A_1 j_{t-1} + u_t$,

$$\operatorname{Var} j_t = A_1 \operatorname{Var} j_{t-1} A_1^T + \Sigma_u.$$

Stationarity implies that $\operatorname{Var} j_t = \operatorname{Var} j_{t-1} = \Gamma_0$, so we have: $\Gamma_0 = A_1 \Gamma_0 A_1^T + \Sigma_u$. We can bring Γ_0 to the left hand side by applying the well-know vec function,

$$\operatorname{vec} \Gamma_0 = (I - A_1 \otimes A_1)^{-1} \operatorname{vec} \Sigma_u,$$

where $\otimes$ is the tensor (Kronecker) product and I is an identity matrix of dimensions $(3B)^2 \times (3B)^2$.