Timing and effects of new advertisement creatives

May 4, 2016

Abstract

In an environment with substantial noise caused by competing brands advertising in close proximity, new creatives provide an opportunity for firms to rejuvenate or refresh campaigns. In this study we document how competing brands launch new creative over time. We then examine evidence of the long run effect of different creatives run for television advertising. Our model builds on past work in marketing on the dynamics of advertising response and extends this framework to a competitive context. A particular feature of our empirical application is that advertising is run at a national level, whilst the response is at a more disaggregate (e.g. city) level. This makes it necessary to look at a hierarchical dynamic model of advertising response. We examine how multiple competitors' decisions to advertise interact with one another and how this translates into each brand's advertising response. These results are used in a dynamic model of advertising response to examine the tradeoff between the cost of implementing new messages, versus reusing older media messages. This work will help guide companies toward more efficient allocation of their marketing resources.

1 Introduction

Marketing organizations spend vast amounts of time and money each year on producing advertising new video based commercials, often referred to in the advertising industry as 'creatives'. The return on investment of producing a new creative hinges on how effective the innovation to the campaigns are on sales response (see e.g.). It has been observed numerous times, however, that over time advertising can wear out, and this can be exacerbated by excess spending (e.g. Naik et al. 1998, Bass et al. 2007, Braun and Moe 2013). Consequently, a substantial proportion of the overall ad budget is required to produce new advertising content. To justify this spending, managers need to quantify how much these creative changes can improve the effectiveness of their advertising spending, and understand when to use old creatives versus when to spend money producing new ones.

This topic has been noted to be of high interest to practitioners. For example, drawing from a quote from Millward Brown (2015):

> Advertisers often ask us how many GRPs they can put behind an ad before it "stops working." They also wonder if past copy can be rerun or if it has no remaining value. These are important financial issues for them. Producing TV ads is expensive and requires a long lead time. Airtime may need to be booked months before actual airing, and an assessment of the number of ads required needs to be made early.

As mentioned in the Millward Brown quote, it is expensive to produce a new video-based creative, especially for TV. Recycling older creatives may work, since how much they wear out can be forgotten or "rejuvenated" (e.g. Naik et al 1998), but such creatives are subject to wear out over time. The upshot is that managers need guidance as to when to draw from the existing stock of creatives, and when to expend resources to replace old advertising creatives, and they need to be able to have this information early enough to execute a change in creatives. As more money is being channelled into digital advertising, a growth in the use of video based ads is making this topic of high practical importance.

Marketing science has provided partial solutions helpful to addressing this managerial issue. For example, Naik, Mantrala and Sawyer (1998) pioneered the use of dynamic effectiveness of ads in a Nerlove Arrow model of brand awareness. Bass, Bruce and colleagues extended this to multiple message types, but do not explicitly consider changes in ad response due to changes in creatives within each appeal. Braun and Moe (2010) were able to examine the introduction of new creatives on existing campaigns, in a digital marketing (banner ads) context and confirm there is significant wear out that drives the need to introduce new creatives. Until recently, all these models downplayed the role of competition. This was modeled more specifically in Naik et al. (2008) in the context of brand awareness among a set of competing brands, but they are missing how creative changes can bring about changes in ad response.

Competition is important, because in mature categories such as the ones in our study, with high penetration rates, brand growth and performance comes largely from attracting buyers from other brands. Advertising is an important tool on this battle ground for market share, and managers should be informed about where their sales come from. The effect of a change in creative must be measured in terms of a brand's ability to draw customers from other brands (i.e. offensive capability), but also in terms its ability to reduce how much a competing brand's advertising can draw from a focal brand (the defensive capability). The converse of this is also important to understand. If a competing brand changes its campaign in some way, how will this impact a focal brand? As is well known (e.g. Ataman et al 2010), such effects take place in both the short and long term, so quantifying the effects of changing creatives must consider time based response.

Accordingly, we develop a dynamic competitive model of brand evolution, in the spirit of Ataman et al. 2008, which includes both short and long term sales response to advertising expenditure. The fundamental goal of the model is to diagnose the effect of a change in creative, and to come up with a forecast of the likely path of a brand's growth if a focal brand were to change its creative. The We adopt the standard approach of allowing advertising to have both short and long term effects, the length of this effect governed by a 'memory' decay parameter that is estimated in the model. As a unique feature of our competitive model, we allow each brand to shaped by the advertising of all competing brands' advertising. While the competing effects of brands have been considered in a more aggregate sense in e.g. Ataman et al. 2008, we consider it explicitly among several of the major brands advertising. A study by Naik et al. 2008 also considers competing brands' advertising, but model only the evolution of awareness, and do not consider dynamics of campaign efficacy, or consider other marketing mix elements. Consistent with past work interested in modeling the effects of brands' marketing mix, in our model, we control for such other promotional events (e.g. product line length, distribution, and promotions such as features and displays), as major drivers of sales over time, and also control for dynamics in how competitive pricing is able to redistribute market shares across the brands. This more complete picture is needed if one were to balance the cost of producing a new creative against the benefits of doing so, namely increase in sales revenue/profit margins.

Another feature of our data is that it is at a city level of aggregation, whereas the majority of the advertising expenditure is at a country level. Accordingly, the dynamics play their role at a national level, but we explicitly model city level variation allowing dynamic parameters to vary across cities.

We highlight a number of important contributions of this work. First, it adds to our understanding about how advertising works, particularly in a dynamic, competitive setting. The most important innovation of this work is the analysis of the market level effects of changing advertising creatives on city level brand performance. In achieving this, we also allow for several other innovations. While the marketing literature has allowed for dynamics of the effectiveness of ad campaigns (e.g. Naik, Mantrala and Sawyer 1998, Braun and Moe 2013), several things are missing. First, doing this in a competitive environment, explicitly recognizing that an ad that is run for one company has the ability to affect other brands. More recent research has recognized that these effects may be positive (spill over effects, e.g. Simester/Anderson). Second, there are no papers that examine how changes in creatives will impact a brand's advertising performance. This is important since the content of advertising is likely to vary over time as brands use more or less effective persuasive appeals. While marketers continue to strive for the most effective appeal in creatives, there is likely to be random variation across appeals. This may also be contextual - e.g. that two leading brands both using highly persuasive appeals but launching them around the same time, could yield cancel one another out. The combination of the two above innovations is also important. This leads to questions about how new creatives actually work. Do they improve the focal brand's own advertising effectiveness, or do they reduce the ability for a competing brand's advertising to draw sales from the focal brand? In terms of spillover, does improving the effectiveness of the focal brand's advertising increase or decrease any existing spillover to another brand?

While our study focuses on television advertising, a natural thought is that much interest is in digital marketing. However, we also higlight that digital marketing campaigns are increasingly using more video based appeals (adapted for the different medium), in favour of static appeals, as more effective ways to engage customers. Thus our insights and methodology are highly relevant for the new media.

1.1 Application and data

Our application focuses on brand level for the top brands within a number of categories of fast moving consumer goods. The categories studied include facial tissues, paper towels, disposable diapers, bathroom tissues and laundry detergents.[1] The data are provided by IRI, and represent 42 of the largest cities in the USA. We have volume and revenue data for each of the UPCs, at a store level. For each city, IRI has a sample of stores for which data is observed. We aggregate the UPC level data to brand, and then aggregate across the stores within each city. Aggregating the data across the 42 cities is problematic in that there is likely to be important city level variation in both outcome (e..g Bronnenberg, Dube and Dhar 2007) and possibly in response variables.

[Table 1 about here.]

Before we examine our data on television advertising, we construct and summarize several other covariates that are known to important in explaining variation in sales. These include measures on *Pricing* where we observe transacted price, in dollars. Given our log-log model specification to construct average prices across stores, we use the geometric mean (suggested by Christen et al 1995). *Promotions* are retailer promotions and include features and displays. Feature promotions represents items featured in price-based advertising by a local retailer. That is, a UPC is designated as on feature if it is included in this advertising, regardless of whether it has a price discount or not. Typically, they include the price discount. We have available both the number of UPCs that were on display in the sample store, as well as whether there was a price discount. *Distribution* is also available, and can be measured in a couple of ways. First, it can be represented by the number of stores sampled, that have this brand available for sale. Second, using a weight according to ACV. We currently use the number of outlets out of the total sampled, as a proxy for the distribution in the city. *Products* The assortment can be measured by the number of UPCs within a

[1]We also have a few extra categories that are more complex, and could be subdivided into smaller sub-categories. These include yogurt, cereals, coffee, and beer. We also have razors, which is complex because of their relation across categories (e.g. blades and handles).

brand. The length of a product line is a measure of the how many different versions (e.g. flavours, sizes) contained in the product line.

Table 1 reports these summary statistics across all stores and cities, and with the unit of time as indicated (typically weekly). The brands represented in these categories are fairly regularly promoted (e.g. see columns labeled fracdnp and fracfnp), and distribution is in general very high, being available across more than four out five stores sampled. The assortment per brand in terms of the number of SKUs is typical of product lines in fast moving consumer goods. The brands include ’private labels’ although advertising explicitly for these brands is not available. There may be other brands that are sold but do not advertise, but our model must be able to include these as competition. They can be affected by brands’ advertising from national brands that do advertise. Note that creative changes can also be a subset of the brands advertised, that in some time slices there may be no brands that introduced new creatives (although for the entire time span, almost all brands that advertise used creative changes).

1.1.1 A deeper dive into the advertising data

TNS provided the data for national level advertising. While these sources were available across different media, such as cable, network, syndicated, and Spanish language channels, we combine all national sources of advertising by summing across advertised dollars. We acknowledge that these advertising sources are a subset of the total amount spent across all other media. Table 2 shows selected summary statistics for each of the categories studied. Each of the brands launches a number of new creatives per year, but this appears to be quite variable across categories. For disposable diapers, the average new creatives launched per year and per brand is 17, but for paper towels this is lower at 6. This appears to be related to the total annual amount spent on advertising per brand. Looking at the amount spent per creative over the entire time they are aired, corroborates this insight. It appears that the average amount spent per creative is around $3-4 million, the maximum being in toilet tissue where it is closer to $5m. Theoretically, if one were to assume that creatives are taken out once they wear out, it appears to be consistent with the notion that wearout is a function of how much cumulative spending there is on the focal creative (e.g. Naik et al 1998).

[Table 2 about here.]

A new creative is identified in our data as any change in creative description, as coded by TNS, that has not previously been recorded.[2] We use data from prior to our observation period, to help us avoid 'initial conditions', so a new creative is one that has not been shown for at least 30 weeks prior to it being used.

Figure 1 shows histograms/density plots for the length of time creatives are used by brands. The histograms are overlaid with estimated densities being Weibull distributions which for most categories appear consistent with simple negative exponential distributions (possibly with the exception of the liquid laundry detergents category).

[Figure 1 about here.]

The table reporting summary statistics (Table 2) and Figure 1 allow us to make several initial observations.

Observation 1. Creatives are used for a relatively short span, with the mean being around 25 weeks (6 months) and many being less than a month in duration.

This observation adds urgency to the need to further understand and measure the effectiveness of new creatives. While there are some category level differences, generally we find the means to be close to 25 weeks. Liquid laundry detergents have particularly short length of run time for creatives, being an average of 19 weeks, and 95% of new creatives are used for less than a year.

The data also allow us to observe the total number of minutes of air time they are run. Generally, the average airtime received for ads on television is around 400 minutes, ranging from around 300 minutes (6

[2] We also used a text based metric for distance of descriptors to reduce the number of creatives to a smaller subset.

hours) for facial tissues, to 480 minutes (8 hours) for toilet tissues. Figure 2 reports histograms for the total spent per creative, which naturally aligns with the total number of minutes.

[Figure 2 about here.]

Observation 2. Creatives are used for a relatively fixed period cumulative period of airtime, before they are removed, with category averages ranging from six to eight hours. The corresponding cumulative amount spent on creatives is generally less than $10 million.

The next observation comes from Table 2, in the column ”fwpct” which measures how much of a brand’s ad spend for that week is accounted for by the new creative launched that week. This gauges how much the initial week is dedicated to any new creative on average. Note that if individual brands tend to launch several creatives in any one week, this would depress this average. Figure 3 also gives a histogram across all creatives introduced to understand the distribution. This indicates that there are some creatives that tend to used exclusively by the brand to replace all previous creatives. However, generally a lot of creatives start of with a low fraction of the budget being spent on them.

Observation 3. For any brand, on average a substantial fraction of the ad budget for that week is spent on a new creative, the first week a creative is run, being 20-30% of any brand’s budget.

[Figure 3 about here.]

[Table 3 about here.]

Table 3 reports the results of regressing the log of length of time (in weeks) that a creative is being used, onto (the log of) several of the variables available in the dataset. In particular we are interested in seeing if the length of time varies considerably by category, and whether the number of properties (channels) and programs used is associated with longer or shorter durations. We also see how much the variation these variables alone seem to account for.

Observation 4. A substantial amount of the variation in the length of time creatives are used, is associated with the way in which the creatives are “trafficked” or placed within media. In particular:

1. Creative appeals are used for longer if they are placed within more programs.
2. They are used for less time if they are on more diverse properties.
3. If there are a larger number of brands, they are used for less time.

It is clear that modeling a large number of creatives, each with its own potential effect could be difficult, especially when advertising effects are typically measured with a great deal of uncertainty. As a proxy we need something that captures key elements of the properties of brands' spending across creatives. Therefore we construct from our advertising data several creative mix elements that we will henceforth use in our analysis.

1. **Number of creatives.** This is the total number of new creatives that any brand uses in a given time period.
2. **Number of creatives, weighted by first week fraction spent.** In the spirit of our observations above, we use the fraction of a brand's ad spend for that week, accounted for by the new creative(s).
3. **Number of creatives, weighted by second week fraction spent.** Same as above, but this is how much is being spent on a second week that the ad is being used.
4. **Novelty.** This being the weighted average age of all creatives being used by a brand. The weight being the dollars spent.
5. **Variety.** This being the weighted average number of creatives being used by a brand for that week. The weight being the expenditure per creative.
6. **Concentration.** The concentration is similar to a Herfindahl index, being the sum of the square of the fraction of the budget accounted for by each creative aired. For example, if a brand is using two creatives, the first one accounting for 0.95 of the budget, the second 0.05, the concentration is $(0.95^2 + 0.05^2) = 0.905$

Different combinations of the above may also be utilized (e.g. the sum of the first and second week fraction spent).

In terms of creative mix, what are the average number of creatived used, the average age and the average concentration? And how much variation is exhibited by these plots? Figure 4 plots the time series for each of these values, and overlays the mean values for each of these on the plots. We see that on average, ads are around 25 weeks old, with the brand ”LUVS” having the oldest creatives. The leading brands, Huggies and Pampers have average ages markedly close together. The average novelty, being the number of creatives being used across all the brand’s advertising within any week, is typically less than 10, but we see the largest brand having up around 17 different ads used in any one week. These figures correspond to the concentration measures, displayed at the bottom.

[Figure 4 about here.]

1.2 Modeling advertising response

We focus here on the combined advertising expenditure for a brand, but acknowledge that a more finessed model would track each creative over time. The following are key aspects we wish to capture in a model of advertising in a competitive environment:

1. Creative changes. Our specific interest is in how the changes in creatives by individual brands can have effects on these brands’ abilities to draw customers in. Further, we examine how they affect the ability for a brand to compete. The key question being, will the addition of a creative be more focused on the brand, or will it improve the brand’s ability to compete?
2. The effectiveness of advertising translates expenditures into a baseline for each brand. We assume that this effectiveness changes over time. For example, wear out of the messages across the brand’s campaigns and across media.
3. Memory decay. Over time, information about a brand is ’forgotten’. So the advertising expenditure today may boost a brand’s appeal, being a state variable, but this cumulative appeal is then assumed to wear out over time at the memory decay rate. We estimate this, but assume it is constant over time, across cities, and across advertised brands.

4. The change in advertising effectiveness (being how much spending augments a brand's value) is also governed explicitly by a change in creative appeal.

5. The effect of other marketing instruments. There are many other aspects that contribute to sales performance, including previous sales/product satisfaction, promotions, advertising not from television etc, product line length, distribution and so on.

6. Competitive effects of other brands' advertising. These could be positive or negative. Creative changes can also affect the ability for competing brands to draw sales from (or to in the case of spillovers) from competing brands.

7. National level expenditure, but city level effects. We focus on network advertising and our sales data are at city levels. We have data on a subset of the total USA, being individual US cities (we have 42 of them).

8. Endogeneity of advertising expenditure, and of the production of new creatives. Since both incur costs, it is essential to understand when these are introduced and how the process by which they are done depends on parameters we are trying to infer.

2 Theoretical development of the dynamic hierarchical effects of advertising

In our application, we take the perspective of a national-level marketing manager observing city level observations of sales and prices, allocating an advertising expenditure over time and across cities. We focus on the effectiveness of network advertising, which as a unit of analysis is observed only at a national level. Expenditure allocation decisions are made on a continuous time basis but we use weekly level sales data so we convert the expenditures to a weekly level. That is, each city is exposed to the same network television patterns. Effectiveness is a dynamic function of the content in the creatives used in the campaign, and how this wears out over time. At any time, there are a number of distinct creatives observed in the market place. The manager can choose to spend on existing/past creatives, or to invest in and launch a new creative. The industry setting is such that there are a total of J competing brands. Not all of them advertise on a specific medium, but all are affected by the brands that do advertise. Furthermore, not all brands change their creatives, but creative changes will affect all brands (including those that do not advertise). In addition to competitive effects and competitive interference effects, the effectiveness of the advertising campaign wears out over time, being linked to the wearout of individual messages (e.g. Naik et al 1998, Bass et al 2008, Braun and Moe 2012).

Several past observations shape our expectations about what to expect in advertising effects and the addition of creatives in a competitive dynamic setting.

2.1 Formal model for the effects of advertising on brands

We start with the standard Nerlove-Arrow type evolution of ”brand” or overall ”advertising” effectiveness (defined as the ability for a dollar spent on advertising to lift sales volume), represented by:

$$\frac{dB_j}{dt} = -\delta B_j + q_{jt} g(A_{jt}) \tag{1}$$

where $g(A_{jt})$ is some transformation of advertising ($= 0$ if $A_{ijt} = 0$). We will return to the issue of including competitive effects of a focal brand’s advertising, multiple media messages and the effect of competing brands’ advertising on the focal brand’s sales.

One way to build dynamic advertising effectiveness is by using a differential equation with respect to time, of brand j’s advertising:

$$\frac{dq_j}{dt} = f(q_j) + \phi E_{jt}$$

where E_{jt} is a counter (general measure?) for the number of new creatives introduced, and $f(q_j)$ is some general function for the evolution of advertising effectiveness, which may be function of recent advertising expenditure patterns.

2.2 Competing brands

We model competition in a flexible way. We are only modeling a subset of national brands that cover much of the market, with some markets having a prominent share private label presence. The sales of brands that do not advertise nationally, must be modeled alongside brands that do advertise nationally (e.g. Private Labels). In this respect, brands that do not advertise are affected by the advertising of competing brands that do.

Let J represent the total number of competing brands modeled. This includes the national brands, any private labels, as well as a 'composite' brand. A subset of this set of brands advertises across the observation series, and the remaining do not advertise (e.g. the PL does not record any specific advertising). Denote J_b be the brands that advertise across the series. Finally, a subset of the J_b brands recorded a change in creative during the period observed. Let this set be denoted by J_{bE} and create a vector of length J_{bE} which records which of the brands that advertised, also changed a creative.

Collectively, we have a Nerlove Arrow model for the effect of advertising on brand $j \in J$ (with $J_{bE} \subseteq J_b \subseteq J$)

$$\frac{dB_j}{dt} = -\delta B_j + \sum_{k=1}^{J_b} q_{jkt} g(A_{kt}) \tag{2}$$

with

$$\frac{dq_{jk}}{dt} = f(q_{jk}) + \phi_{jk} E_{kt}$$

representing the effect of brand $k \subseteq J_{bE}$ adding $E_{kt} > 0$ creatives at time t.[3]

[3]Note that an interesting model may be to allow for creatives to have an effect on other brands' effects, if that makes any sense and is identified:

$$\frac{dq_{jk}}{dt} = f(q_{jk}) + \sum_{k=1}^{J_{bE}} \phi_{jk} E_{kt}$$

3 Two-level hierarchical model

Assume we have outcome (e.g sales) data for J brands, in N cities, and observe these data over T time periods. We represent this outcome (Y_t) at time t as a $N \times J$ matrix, with the rows representing city level observations, and columns representing brand sales (outcome) data. We have a hierachy of city level at the lowest (each city has its own sales) and at the highest level we have dynamics at the mean level (e.g. mean prices, or the effects of national level advertising). This is written as:

$$Y_t = F_{11t}\Theta_{11t} + F_{12t}\Theta_{12} + v_{1t} \quad (3)$$

$$\Theta_{11t} = F_{2t}\Theta_{2t} + v_{2t} \quad (4)$$

$$\Theta_{2t} = \tilde{G}_t\Theta_{2,t-1} + \tilde{H}_t + w_t \quad (5)$$

The components that affect each city's sales directly, are in the F_{12t} matrix, with a corresponding non-time varying coefficient matrix. The time varying component at the city level is contained in the $F_{11t}\Theta_{11t}$ component. In addition we have an innovation function (sometimes called a control variable) in the evolution equation, H_t. This component shifts elements of Θ_{2t} but is not relative to it.

We use a matrix normal distribution for all covariance terms:

$$v_{lt} \mid \Sigma, V_l \sim N(0, V_l, \Sigma) \quad (6)$$

$$w_t \mid \Sigma, W \sim N(0, W, \Sigma) \quad (7)$$

Each matrix normal distribution has a left and right variance matrix, e.g. V_1, Σ respectively. The right variance governs (column) cross equation covariatance. The left variance captures row covariance, which is either concurrent (V_l) or time based variation (W). The left variance V_1 represents variation across cities. The left variance V_2 represents concurrent variance across mean values for different state variables. We can simplify notation a bit by using a set $\Psi = \{V_1, V_2, W\}$.

At the first level we have $\bar{Y}_t = Y_t - F_{12t}\Theta_{12}$, which does not have a hierarchical counterpart (i.e.

homogenous response to covariates contained in F_{12t}). The Θ_{11t} component then has both time varying and non-time varying heterogeneous responses. We can rewrite our HDLM as:

$$\bar{Y}_t = F_{11t}\Theta_{11t} + v_{1t} \tag{8}$$

$$\Theta_{11t} = F_{2t}\Theta_{2t} + v_{2t} \tag{9}$$

$$\Theta_{2t} = \tilde{G}_t\Theta_{2,t-1} + \tilde{H}_t + w_t \tag{10}$$

We use the tilde ($\tilde{\cdot}$) in the above to represent intermediate variables (those that depend on other parameters).

In the full competitive model, all covariates (advertising, price and promotions) have both an own and cross effect. In the matrix normal set up of the HDLM above, this is automatically specified by having a matrix normal of the brand sales in the columns, and the covariates are each brands' covariates.

3.1 Targeting of creatives

The introduction of creatives can be modeled as a poisson regression. There are two models here. First is a model that is just the introduction of the creative (0/1) and the number introduced at any time.

The second involves the removal or reduced emphasis on the incumbent creatives. The joint of the two distributions is here of interest.

3.1.1 Observable candidates

Candidates for observable components can be quite broad. Here we identify two components. The first being the average age of the creatives currently being used. The second being the number of creatives being used by the focal brand. Other candidates can be tested. We stack these variables in the matrix **CM** being of dimension $T \times R$, with R being the number of variables times the number of brands advertising.

3.1.2 Unobservable candidates

Unobservable targeting variables are the response rates of sales to advertising. They vary over time, and at time t they are identified based on the information available in M_{2t}.

3.1.3 PDF of Poisson conditional on data and latent effectiveness

We allow for a Poisson process to guide the number of creatives introduced at any time point, with the number of creatives being equal to some integer $E_t = \{0, 1, 2, \ldots\}$.

The pmf is:

$$p(E_1, E_2, \ldots, E, \ldots, E_T | CM, \gamma^E, \gamma^q, \Theta_2) = \prod_{j=1}^{J} \prod_{t=1}^{T} \frac{\lambda_{jt}^{E_{jt}} \exp^{-\lambda_{jt}}}{E_{jt}!} \tag{11}$$

where λ_{jt} is given a link function:

$$\log(\lambda_{jt}) = CM_t \gamma_j^E + \Theta_{2t,(1+j,\cdot)} \gamma_j^q$$

$\Theta_{2t,(1+j,\cdot)}$ being the row of the national level state matrix at time t corresponding to what the advertising of that brand does to the focal brand, and all other brands, respectively. That is, the addition of a creative by brand j takes into account the impact of adding that creative onto the own effectiveness of that brand q_{jjt} as well as effects of adding that creative onto the ability for that brand to draw sales from other brand (being $q_{kjt}, \forall k = 1, \ldots, J, k \neq j$). The parameter vectors $\{\gamma_j^E, \gamma_j^q\}$ are respectively of dimension $R \times 1$ and $J \times 1$. These are then stacked as matrixes so that $\gamma^l = \{\gamma_1^l, \ldots, \gamma_j^l, \ldots, \gamma_J^l\}, \forall l \in \{E, q\}$.

An important practical question here is how much the observables explain the choice of whether to add a creative or not.

3.2 Endogeneity of advertising expenditure

The amount of advertising allocated to each week is also endogenous. Given the tactical nature of this decision we do not worry here about the annual amount of advertising. However, we do worry about the allocation of advertising expenditure over time. Again, the observables can include several variables (to be tested), and we will stack these as CA.

Hypothesis being, that they would be negatively driven by brand equity. Lower the brand equity, higher the advertising expenditure.

A couple of approaches could be used here. Possibly the simplest is to model this as a censored normal, the lower bound being zero. The expenditure then becomes:

$$P(A_1, \ldots, A_t, \ldots A_T|\mu, \sigma) = \prod_{t=1}^{T}\prod_{j=1}^{J}\frac{1}{\sigma}\phi\left(\frac{A_{jt}-\mu_{jt}}{\sigma}\right)\left[(1-\Phi\left(\frac{A_{jt}-\mu_{jt}}{\sigma}\right)\right]^{-1} \tag{12}$$

with μ being a time series matrix of dimension $T \times J$, each element μ_{tj} being a function of observable and unobservables:

$$\mu_{tj} = \Theta_{2t,j\cdot}\gamma^{B} + CA_t\gamma^{CA}$$

For both advertising expenditure and new creatives, the process is conditional on the posterior value for the parameters in Θ_{2t}. We will return to how to calculate this once we talk about the main model to infer the states.

Data likelihood

We write the data likelihood, recognising that simple substitutions can be made to take care of the time invariant and homogenous parameters. Let D_{t-1} represent all information (including state variables) available at time t. So D_0 is initial information about the states and priors. At any time period, $\Sigma \mid D_{t-1}$ is distributed as an Inverse Wishart ($IW(\nu_{t-1}, \Omega_{t-1})$. For any time period, the joint density of the data Y_t and

Σ is a matrix normal inverse Wishart (or a product of a matrix normal with inverse Wishart):

$$P(\bar{Y},\Sigma \mid \mathcal{D}_0,\Psi) = \prod_{t=1}^{T} P(\bar{Y}_t \mid \Sigma, D_{t-1},\Psi) P(\Sigma \mid D_{t-1},\Psi) \tag{13}$$

$$= \prod_{t=1}^{T} (2\pi)^{-\frac{NJ}{2}} |Q_t|^{-\frac{J}{2}} |\Sigma|^{-\frac{N}{2}} \exp\left[-\frac{1}{2}\operatorname{tr}\left(\left(\bar{Y}_t - f_t\right)' Q_t^{-1}\left(\bar{Y}_t - f_t\right)\Sigma^{-1}\right)\right]$$

$$\times IW(\nu_{t-1},\Omega_{t-1}) \tag{14}$$

Integrating out Σ (see our technical appendix on matrix T) gives the following data likelihood:

$$P(\bar{Y}|\cdot) = \prod_{t=1}^{T} P(\bar{Y}_t | y_{1:t-1},\cdot)$$

$$= \mathcal{K}\left(\prod_{t=1}^{T} |Q_t|^{-\frac{J}{2}}\right) |\Omega_0 + \sum_{t=1}^{T} \left(\bar{Y}_t - f_t\right)' Q_t^{-1}\left(\bar{Y}_t - f_t\right)|^{-\frac{\nu_0 + TN}{2}} \tag{15}$$

where:

$$\mathcal{K} = \pi^{-\frac{NJT}{2}} \frac{\Gamma_J\left(\frac{\nu_0+TN}{2}\right)}{\Gamma_J\left(\frac{\nu_0}{2}\right)} |\Omega_0|^{-\frac{\nu_0}{2}}$$

3.3 The conditional distribution of Θ_{2t}

First note that, for some $\Sigma_t \sim IW(\nu_t,\Omega_t)$:

$$p(\Theta_{2t}|\Theta_{2,t+1},\ldots,\Theta_{1,t}) \sim N(U_t^*, u_t^*, \Sigma_t)$$

where $U_t^* = \left(C_t^{-1} + G_t' W^{-1} G_t\right)^{-1}$, and $u_t^* = U_t^*\left(C_t^{-1} M_t + G_{t+1}' W^{-1}\Theta_{2,t+1}\right)$. The component Σ_t^* is the estimated value of the common right covariance using data available up to time period t. Note in the recursion that the last period we observe, and the one that we assume here that advertising is set according to, is at time t, which is given simply by $\Theta_{2t} \sim N(M_t, C_t, \Sigma_t)$, with Σ_t being $IW(\nu_t,\Omega_t$ and $\nu_t = \nu_0 + t \times N$, and $\Omega_t = \Omega_0 + \mathcal{A}_t$ and $\mathcal{A}_t = \sum_{\tau=1}^{t}(Y_\tau - f_\tau)' Q_\tau^{-1}(Y_\tau - f_\tau)$.

In the algorithm above, the likelihood must include the joint posterior of both the likelihood of the data in Y_t conditional on the state space system, but also the endogenous likelihood contributions due to the specific components of Θ_{2t} that are being used to set advertising expenditure and creative additions. We describe this in a bit more detail in the appendix, when we describe the recursion over time. Intuitively, the national brand manager can infer Θ_{2t} at time t using all available information available only up to time t.

This component of the likelihood at time t is then the product of the time t likelihood, and the normal distribution over Θ_{2t}:

$$\begin{aligned} P(A_t|\mu_{\cdot t},\sigma) \times P(\Theta_{2t} \mid M_t, C_t, \Sigma_t) &\propto \prod_{j=1}^{J} \frac{1}{\sigma}\phi\left(\frac{A_{jt}-\mu_{jt}}{\sigma}\right)\left[(1-\Phi\left(\frac{A_{jt}-\mu_{jt}}{\sigma}\right)\right]^{-1} \\ &\times \quad |C_t|^{-J/2}|\Sigma_t|^{-N/2}\exp\left\{-tr\left[(\Theta_{2t}-M_t)'C_t^{-1}(\Theta_{2t}-M_t)\Sigma_t^{-1}\right]\right\} \end{aligned} \tag{16}$$

and for the creatives:

$$\begin{aligned} P(E_t|CM_t,\gamma^E,\gamma^q,\Theta_{2t}) \quad &\times \quad P(\Theta_{2t} \mid M_t, C_t, \Sigma_t) \propto \prod_{j=1}^{J} \frac{\lambda_{jt}^{E_{jt}} \exp^{-\lambda_{jt}}}{E_{jt}!} \\ &\times \quad |C_t|^{-J/2}|\Sigma_t|^{-N/2}\exp\left\{-tr\left[(\Theta_{2t}-M_t)'C_t^{-1}(\Theta_{2t}-M_t)\Sigma_t^{-1}\right]\right\} \end{aligned} \tag{17}$$

(I am not sure if I'm double counting the likelihood contribution given by the $N(M_t, C_t, \Sigma_t)$ component here, but I suspect that I am. Perhaps we can use the expected value of Θ_{2t} being M_{2t} and avoid altogether the rest?

3.4 Specifying Θ_{2t}

The matrix Θ_{2t} is a (dense) matrix of time varying parameters (also called state variables). Without the $P \times J$ matrix of time varying parameters, the rows of this correspond to

$$\Theta_{2t} = \begin{bmatrix} B_{1t} & B_{2t} & \dots & B_{Jt} \\ q_{11t} & q_{21t} & \cdots & q_{J1t} \\ q_{12t} & q_{22t} & \cdots & q_{J2t} \\ \vdots & & & \\ q_{1J_bt} & q_{2J_bt} & \cdots & q_{JJ_bt} \end{bmatrix} \tag{18}$$

with F_{2t} being the matrix (dimension $N \times (J_b + 1)$) that translates these states to the city level. For example, corresponding to the above:

$$F_{2t} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \vdots & & & \\ 1 & 0 & \dots & 0 \end{bmatrix} \tag{19}$$

being a column of ones, followed by J_b columns of zeros. The zero elements of this matrix make the ad effectiveness parameters/states latent, with their role only being played in the evolution matrix below. For (18) and (19) they combine to provide a $N \times J$ matrix which just depends on the intercept (B_{jt} in 1).

3.4.1 Adding time varying effects of other covariates

To add time varying effects of other covariates, we add in rows corresponding to each effect to Θ_{2t}, e.g. adding J rows[4], a price "mean" parameter for each price variable (i.e. this is the mean value for the price

[4]This is a bit confusing, but because we have cross effects for each covariate we add, then adding in just price adds $P = J$ covariates. If we allow two types of covariates (e.g. price and promotion) then $P = J + J$

elasticity across cities):

$$\Theta_{2t} = \begin{bmatrix} B_{1t} & B_{2t} & \dots & B_{Jt} \\ q_{11t} & q_{21t} & \dots & q_{J1t} \\ q_{12t} & q_{22t} & \dots & q_{J2t} \\ \vdots & & & \\ q_{1J_bt} & q_{2J_bt} & \dots & q_{JJ_bt} \\ \theta^p_{11t} & \theta^p_{21t} & \dots & \theta^p_{J1t} \\ \theta^p_{12t} & \theta^p_{22t} & \dots & \theta^p_{J2t} \\ \vdots & & & \\ \theta^p_{1Jt} & \theta^p_{21t} & \dots & \theta^p_{JJt} \end{bmatrix} \tag{20}$$

Then F_{2t} becomes (with dimension $N(1+P) \times (1+J_b+P)$):

$$F_{2t} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ \vdots & & & & & & & \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ \vdots & & & & & & & \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{21}$$

3.5 Specifying F_{11t}

The city level matrix of F_{11t} is the mean for the distribution of $\bar{Y}_t$ as a (simple additive) function of the covariates (and latent space) at the city level. Without any additional time varying effects of covariates, this is an $N \times N$ identity matrix:

$$F_{11t} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & 0 & \ddots & \\ 0 & 0 & \dots & 1 \end{bmatrix} \tag{22}$$

The addition of city specific price covariates, for which the effects vary over time, then the matrix is $N \times N(1+P)$ where P corresponds to the number of covariates (times J). For example, consider adding price for J brands, we will add $P = J$ covariates to F_{11t}:

$$F_{11t} = \begin{bmatrix} 1 & p_{11t} & p_{1jt} & \dots & p_{1Jt} \dots & 0 & & & & & \dots & 0 \\ \vdots & & & & & & & & & & & \\ 0 & 0 & 0 & 0 & \dots & 1 & p_{n1t} & p_{njt} & \dots & p_{nJt} & \dots & 0 \\ \vdots & 0 & \ddots & & & & & & & & & \end{bmatrix} \tag{23}$$

where p_{njt} is price in city n for brand j at time t. In the above, the dimension would be $N \times N(1+P)$.

3.6 Specifying $\tilde{G}_t$

The corresponding ”evolution” matrix $\tilde{G}_t$ is $(1 + J_b + P) \times (1 + J_b + P)$, and is upper triangular. We illustrate this below by ignoring the $\tilde{G}_t$ component for the P time varying components (which would just be an identity matrix of dimension $P \times P$). Let $\tilde{g}(A_{jt})$ be some transformation function of ad spend for brand j at time t.

$$
\tilde{G}_t = \begin{bmatrix} (1-\delta) & \tilde{g}(A_{1t}) & \dots & \tilde{g}(A_{Jbt}) \\ 0 & 1 & \dots & 0 \\ \vdots & 0 & \ddots & \\ 0 & 0 & & 1 \end{bmatrix} \tag{24}
$$

Computing H_t

The innovation component adds an amount to each state parameter, and for each brand. Therefore the dimension of the matrix $\tilde{H}_t$ is $(1 + J_b + P) \times J$. Again ignoring any additional time varying effects of covariates (so $P = 0$) we have:

$$
H_t = \begin{bmatrix} 0 & \dots & 0 & \dots & 0 \\ \phi_{11}E_{1t} & \phi_{21}E_{1t} & \dots & & \phi_{J1}E_{1t} \\ \vdots & \ddots & & & \\ \phi_{1j}E_{jt} & \dots & \phi_{jj}E_{jt} & \dots & \phi_{Jj}E_{jt} \\ \vdots & & & & \\ 0 & \ddots & & & \\ \vdots & & & & \\ \phi_{1J_{bE}}E_{J_{bE}t} & \phi_{2J_{bE}}E_{J_{bE}t} & \dots & & \phi_{JJ_{bE}}E_{J_{bE}t} \end{bmatrix} \tag{25}
$$

To correspond with Θ_{2t}, there will be an additional P rows to $\tilde{H}_t$ corresponding to additional time varying covariates. Note that some of the rows (if $J_{bE} \subset J_b$) will be zeros, if no new creatives are introduced for a particular brand that advertised over some time frame.

Priors

We need to choose initial values for M_{20} and C_{20}. What do these matrices mean, intuitively? They indicate prior information about the starting states (the means across cities). This could come from theory, another process, or could be made quite diffuse. For example, the way we specify the ad effectiveness (above) could provide us with some prior on the initial state that is constrained to be close to 0. Similary, our understanding of price sensitivity is around -1 to -2 so we could provide such a information as a prior. Of particular interest is the "new" creative effectiveness. The overall effect of this should be somewhat proportional to the effect of the base campaign.

Prior on δ: if δ really is between 0 and 1, we could make the prior uniform. But is that realistic? We probably want a density that places zero probability at 0 and 1, or at least 0. Does the literature give us any prior information about what this decay parameter should be? Yes, it says that δ is usually around $0.1 - 0.2$ (for weekly data). We could use a beta$(1,3)$ as a prior. One problem this may raise (in the GDS) is that the parameter is constrained between zero and one.

Priors on $c_{1:J}$ and $u_{1:J}$: depends on the domain. Are they all between 0 and 1. Also, could they be correlated? Would it make sense that if one brand had high wearout, another brand could as well? Do we have prior information on this? Again, c depends on the ad effectiveness value, but is unlikely to be much different from 0. The w value depends on the scale used for advertising expenditure and is expected to be positive. The ϕ parameters are likely to be small since it is unlikely any one creative can have a substantial impact on the overall effectiveness of the campaign.

Priors on V_1, V_2 and W: first, we need better intuition about what these matrices represent. Then, we can come up with a range of reasonable values for the parameters. Given the complexity of the model, we will need to regularize it with prior information. And it would be good to give these priors careful though. Way too many marketers are careless with their "uninformative" priors.

3.7 Prediction

Once we estimate these top-level parameters, we might want to simulate data. That means we need posterior predictive distributions of Y. Can we do that without simulating the Θ parameters directly? Note that we do not collect any Θ draws during the estimation process, since they are all integrated out. We should be able to use forward filtering, backward sampling (possibly with smoothing) to obtain the city level state variables.

References

Ataman, M. Berk, Carl F. Mela, Harald J. van Heerde. 2008. Building brands. *Marketing Science* **27**(6) 1036–1054. doi:10.1287/mksc.1080.0358. URL `http://mktsci.journal.informs.org/cgi/content/abstract/27/6/1036`.

Bass, Frank M., Norris Bruce, Sumit Majumdar, B.P.S Murthi. 2007. Wearout effects of different advertising themes: a dynamic bayesian model of the advertising-sales relationship. *Marketing Science* **26**(2) 179–195.

Braun, Michael, Wendy W Moe. 2013. Online display advertising: Modeling the effects of multiple creatives and individual impression histories. *Marketing science* **32**(5) 753–767.

Naik, Prasad A., Murali K. Mantrala, Alan G. Sawyer. 1998. Planning media schedules in the presence of dynamic advertising quality. *Marketing Science* **17**(3) 214–235.

Naik, Prasad A, Ashutosh Prasad, Suresh P Sethi. 2008. Building brand awareness in dynamic oligopoly markets. *Management Science* **54**(1) 129–138.

Iterative estimation

To estimate the likelihood in Equation (15), we need to compute $\bar{Y}_t$, f_t, and Q_t. The matrix Y_t is the observed, dependent variable, so we need to get f_t and Q_t, and $\bar{Y}_t = Y_t - F_{12t}\Theta_{12}$. Conditional on estimates from time $t-1$, and all data and prior information, we can follow the following algorithm at time t.

1. Compute $\tilde{G}_t$ using A_t, c, u, ϕ and δ, then $\tilde{G}_t$ to include any additional time varying effects for covariates. Similarly, create $\tilde{H}_t$ including $P \times J$ matrix of zeros.
2. Set $a_{2t} = \tilde{G}_t M_{2,t-1} + \tilde{H}_t$.
3. Set $f_t = F_{11t}F_{2t}a_{2t}$.
4. Set $R_{2t} = \tilde{G}_t C_{2,t-1} \tilde{G}_t' + W$
5. Set $R_{1t} = F_{2t}R_{2t}F_{2t}' + V_2$
6. Set $Q_t = F_{11t}R_{1t}F_{11t}' + V_1$
7. Set $S_{2t} = R_{2t}\left[F_{11t}F_{2t}\right]'$
8. Set $M_{2t} = a_{2t} + S_{2t}Q_t^{-1}(\bar{Y}_t - f_t)$
9. Set $C_{2t} = R_{2t} - S_{2t}Q_t^{-1}S_{2t}'$

Then, iterate over t to estimate the data likelihood. Note that the homogenous time invariant component at level 1 of the hierarchy is handled by the transformed variable, $\bar{Y}_t$ which appears in the posterior and can be numerically estimated.

Within this algorithm, we also include the likelihood for advertising expenditure and decision to add creatives, being conditional on the state of Θ_{2t}. Thus we add to the log likelihood at each time period, the log of Equations (12), but we need to have the conditional distribution of Θ_{2t} at time point t being calculated on all available information to that date and employing backward sampling. So, this proceeds as discussed in Landim and Gamerman (2000), page 39. Fortunately, in our model we simply allow advertising to be determined on the basis of the last observed data point, which means we do not need to run this algorithm backward for all $\tau = 1 \text{to} t$, and all values needed in the likelihood are available at time point t directly in the recursion itself.

Data structures

The matrices A ($T \times J_b$) and F_{12t} are standard, dense covariate structures. The matrix F_{11t} is given above, and is sparse. Similary with F_{2t}. We separated out the time-invariant homogenous effects in $F_{12t}\Theta_{12}$, so the matrix F_{12t} is $N \times K$. These are non hierarchical and non time varying in their effects.

Specifying parameters

Table 4 summarizes the parameters that need to be estimated, assuming that the covariance matrices are stationary. The number of parameters lists is the number of *unique* elements. For example, in a symmetric matrix there are, at most, $k(k+1)/2$ unique parameters. But this can still be a large number, so we should think about some kind of dimensionality reduction.

[Table 4 about here.]

To estimate these parameters, we should transform them all to be unbounded. Otherwise, we need to modify the GDS algorithm to handle constrained optimization and simulation (which is possible, but tedious and uninteresting).

For the dense cases of V_1, V_2 and W, typically we would estimate the elements of the lower Cholesky decomposition (taking logs of the diagonal elements to ensure that they are positive). If we add structure to those matrices, we need to reconsider the transformation. However, block diagonals should still allow us to use the Cholesky decomposition approach.

.1 Data summaries

[Table 5 about here.]

[Table 6 about here.]

.2 Other data issues

- We are using 42 markets that overlap with IRI/TNS. The IRI dataset is supposed to cover 50 markets (some are excluded because of high concentration which makes retailers easy to identify). That means we are missing a further 8 IRI markets which TNS has not adequately covered. These include some smaller cities, but also 'cities' labeled as states. Accordingly, the markets we have:

Market Name:	Number of stores:	
ATLANTA	295	BIRMINGHAM/MONTG.
253		
BOSTON	351	CHARLOTTE
265		
CHICAGO	580	CLEVELAND
125		
DALLAS, TX	419	DES MOINES
77		
DETROIT	313	GRAND RAPIDS
100		
GREEN BAY	77	HARRISBURG/SCRANT
278		
HARTFORD	235	HOUSTON
305		
INDIANAPOLIS	146	KANSAS CITY
158		
KNOXVILLE	147	LOS ANGELES
854		
MILWAUKEE	204	MINNEAPOLIS/ST. PAUL
144		
NEW ORLEANS, LA	213	NEW YORK
903		
OKLAHOMA CITY	74	OMAHA
121		
PHILADELPHIA	388	PHOENIX, AZ
309		
PORTLAND,OR	230	PROVIDENCE,RI
91		
RALEIGH/DURHAM	312	RICHMOND/NORFOLK
257		
ROANOKE	233	SACRAMENTO
217		
SALT LAKE CITY	100	SAN DIEGO
283		
SAN FRANCISCO	364	SEATTLE/TACOMA

- We deal separately with Liquid and Powder laundry detergents. However, there are several brands that span the two categories, meaning they may have some spillover.
- In Liquid Laundry Detergents, Xtra brand has no reported advertising by TNS. It is owned by Church & Dwight, makers of Arm & Hammer. It is a low value brand, bought by Church & Dwight in 2001. Average price point over the data is around 1/3 of the premium national brands, and substantially lower than private label. Similarly, AJAX is quite high in share, but no TV advertising is observed.
- TV Advertising is all we collected - but could there be other forms of advertising we need to study for clutter (e.g. magazine, print, billboard)? For TV advertising - spot TV is regional, but other forms (e.g. Cable TV, SLN, Network) are national only. I would assume that the national advertising is observable in each market. However, how can we add up spot with national? Or should we keep them separate?
- We have no ratings for this data, but we do have expenditures.
- There are around 800 ”properties” for advertising on TV. Perhaps we can collect data on these properties to identify similarity here - http://www.globalcommnet.com/comgrp.htm gives databases on these. They are around $1300 to purchase so we need to think about whether this is needed. There's also the issue of resolution of the code names to TV channels - e.g. is ”AFAM” the channel known as ”ABC Family”?
- Combining the data with laundry detergents sales and advertising, we have a decomposition of ad dollars into ”national” versus ”spot”, the latter being only targeted at individual TV stations. We measure both, but the ad $ for national TV advertising is obviously of broader scope than that for spot, so the two cannot be easily aggregated.
- The excel spreadsheet ”summaries.xlsx”, (see sheet lld), gives the overall (aggregated over 313 weeks) data. Note that we need to probably throw out data prior to 2002 because of the lack of TNS coverage of this data. We do see that the top brands advertise. Some brands advertise only spot. We see that Tide has close to 50% of its revenues on advertising but note that this is because national advertising supports all US, whereas sales are measured only for the 42 markets for which data was collected.
- Aggregation of promotions: we weight any aggregation for features by volume sold on each feature. They are coded fvol1, fvol2, etc. for different types of features. Not all features are used (or measured?) in the database. If we want to aggregate further, we need to do so by correctly aggregating by volume at the UPC level. This means rerunning it from the beginning.
- Aggregation of advertising: we simply add up advertising across time and markets. While this gives us a ”correct” aggregation for purposes of identifying how much was spent on each type of advertising, it is not correct from the perspective of estimating lag structures. For this, we would need to take into account data interval issues (e.g. see Tellis and Franses 2006). The upshot is that we would need to account for the ”unit exposure time”, which may well be measured in hours.

category	Nbrand	revenue	price	fracdnp	fracfnp	fracdist	num products	natadspend
dpp	4	1486314.6	0.25	0.04	0.26	0.84	34	784207
ptw	6	2246319.9	1.65	0.11	0.13	0.84	15	571483
fti	3	941035.7	1.02	0.10	0.23	0.90	20	560390
tti	4	2429413.6	0.46	0.12	0.18	0.89	14	541805
lld	8	2143991.6	0.77	0.13	0.17	0.80	15	627614

Table 1: Broad summary statistics for all 42 cities in sample, and for 226 weeks of data.Categories being lld = liquid laundry detergents, tti = toilet tissue, fti = facial tissue, dpp = disposable diapers, ptw = paper towels. Revenues are weekly market revenues and represents a sample of stores and is aggregated across 42 cities. Price represents average price per volume. Natadspend is weekly expenditure on national advertising by all brands, fracdnp/fnp is the fraction of UPCs available that were on display/feature, fracdist = average distribution (number of outlets), numproducts is the average number of SKUs per brand.

categoryname	avgdolpa	totdol	fwpct	minutes	Nc	ml	p25l	p50l	p75l	p95l
dpp	57.95	3.34	0.19	408	17.3	28	8	19	35	88
ptw	22.73	3.83	0.27	402	5.9	24	7	20	35	62
fti	32.05	3.06	0.29	297	10.5	24	7	16	26	64
tti	42.10	4.82	0.28	483	8.7	27	6	22	37	78
lld	23.62	3.45	0.30	347	6.8	19	7	15	29	52

Table 2: Advertising and creative summary statistics per category, with avgdolpa = average annual advertising expenditure per brand, totdol = total amount spent on new creative over its lifetime, fwpct = fraction of a brands ad budget spent on new creative in first week of showing, minutes = total minutes of ad exposure per new creative, Nc = number of new creatives per brand per year, ml = average length of lifetime for creative, and pXXl being the (XX= 25,50, 75 and 95) percentiles for length of time creatives are used.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.8604	0.1148	7.49	0.0000
log(nproperty)	-0.3378	0.0433	-7.81	0.0000
log(nprograms)	0.6989	0.0477	14.64	0.0000
log(nprogtype)	-0.0931	0.0812	-1.15	0.2522
log(nbrands)	-0.2058	0.0637	-3.23	0.0013

Residual standard error: 0.7611 on 763 degrees of freedom Multiple R-squared: 0.5847, Adjusted R-squared: 0.5825 F-statistic: 268.5 on 4 and 763 DF, p-value: $<$ 2.2e-16

Table 3: OLS regression of (log) time a creative is used onto several of the variables available from the advertising data. nproperty is the number of properties being used (being a specific channel). nprograms is the total number of programs across all properties, and the nprogtype variable is the counterpart for types or genres of program (e.g. comedy or sports). Category codes: fti = facial tissues, tti = toilet tissue, lld = liquid laundry detergent, ptw = paper towels, dpp (base category) = disposable diapers.

Symbol	Note	Num Pars (if dense)	Reduce by...	reduced parameters
V_1	symmetric pos-def	$N(N+1)/2$	make diagonal	N
			spatial structure	something $> N$
V_2	symmetric pos-def	$[N^2(1+P)^2 + N(1+P)]/2$	block diagonal?	$N(2+P(3+P))/2$
W	symmetric pos-def	$(1+J_b+P)(2+J_b+P)/2$	diagonal	$(1+J_b+P)$
			block diagonal	$1+J_b(J_b+1)/2+P(P+1)/2$
δ	scalar between 0 and 1	1		
ϕ	dense matrix	$J_{bE} \times J$	make symmetric	$J_{bE}(J_{bE}+1)/2$
Θ_{12}	time invariant homogenous coefficient matrix	$K \times J$	no intercept	

Table 4: Parameters to be estimated

Brand	Dollars	Volume	Dist	National Ad $	Spot Ad $	Market share	Cum Share
Toilet Tissue:							
CHARMIN	$250,763,840	533,030,249	313	$214,667,239	$17,553,844	22.9%	22.9%
QUILTED NORTHERN	$173,645,710	389,624,197	313	$58,925,560	$2,500,336	15.9%	38.7%
PRIVATE LABEL	$165,616,005	479,407,017	313	$-	$-	15.1%	53.9%
KLEENEX	$163,294,811	338,841,325	313	$-	$-	14.9%	68.8%
SCOTT	$159,951,512	254,300,918	313	$29,078,317	$523,714	14.6%	83.4%
ANGEL SOFT	$105,089,175	310,113,901	313	$18,195,027	$4,780,763	9.6%	93.0%
MARCAL	$23,886,389	53,509,233	313	$-	$-	2.2%	95.1%
SOFT N GENTLE	$21,192,702	84,094,494	313	$-	$-	1.9%	97.1%
MD	$21,078,627	68,080,408	313	$-	$-	1.9%	99.0%
SOFT WEVE	$4,341,678	7,995,797	238	$-	$-	0.4%	99.4%
Paper Towels:							
BOUNTY	$271,560,954	142,207,581.2	313	$252,457,770	$21,640,133	37.4%	37.4%
PRIVATE LABEL	$148,956,234	119,085,327.4	313	$-	$-	20.5%	57.9%
BRAWNY	$80,835,038	52,832,260.07	313	$56,505,822	$2,276,223	11.1%	69.0%
SCOTT	$76,726,399	48,671,380.82	313	$17,127,598	$70,423	10.6%	79.5%
VIVA	$59,218,219	19,512,923.54	313	$1,926,090	$4,964,158	8.1%	87.7%
SPARKLE	$47,773,828	34,972,398.62	313	$8,811,822	$32,325	6.6%	94.3%
MARCAL	$18,967,621	14,502,229.53	313	$-	$-	2.6%	96.9%
MARDI GRAS	$11,360,747	7,388,144.459	313	$-	$-	1.6%	98.4%
SO DRI	$3,861,419	2,914,336.05	313	$-	$-	0.5%	99.0%
CORONET	$1,861,167	1,667,048.728	218	$-	$-	0.3%	99.2%
Laundry Detergents:							
TIDE	$307,306,050	299,505,522	313	$247,548,819	$13,494,563	39.0%	39.0%
ALL	$82,954,250	111,199,555	313	$60,716,782	$196,701	10.5%	49.5%
PUREX	$67,518,897	138,550,398	313	$6,415,529	$9,122	8.6%	58.1%
WISK	$48,014,963	53,141,762	313	$-	$2,939,501	6.1%	64.2%
ARM & HAMMER	$45,589,355	93,379,676	313	$-	$36,768	5.8%	70.0%
GAIN	$37,833,065	45,265,242	313	$101,721,781	$6,124,805	4.8%	74.8%
CHEER	$31,566,933	29,513,140	313	$47,910,200	$5,794,025	4.0%	78.8%
XTRA	$31,319,650	92,634,914	313	$-	$-	4.0%	82.8%
PRIVATE LABEL	$25,332,762	55,358,084	313	$-	$-	3.2%	86.0%

Brand	Dollars	Volume	Dist	National Ad $	Spot Ad $	Market share	Cum Share
Facial Tissues:							
KLEENEX	$173,025,002	159,505,663	313	93,057,178	$574,407	50.6%	50.6%
PUFFS	$62,897,664	53,416,278	313	83,266,867	$6,701,733	18.4%	69.0%
PRIVATE LABEL	$60,369,968	79,315,420	313	$-	$-	17.7%	86.7%
SCOTTIES	$35,955,388	45,432,198	313	$-	$-	10.5%	97.2%
MARCAL	$3,910,642	7,233,379	313	$-	$-	1.1%	98.4%
MARCAL FLUFF OUT	$1,256,932	1,216,116	313	$-	$-	0.4%	98.7%
SOFT N GENTLE	$1,206,571	1,566,427	313	$-	$-	0.4%	99.1%
ELLIAIR	$1,161,514	2,009,378	225	$-	$-	0.3%	99.4%
SOFITELLE	$374,144	698,884	296	$-	$-	0.1%	99.5%
NOBRAND	$342,019	269,518	313	$-	$-	0.1%	99.6%
SILKY TOUCH	$268,192	343,474	313	$-	$-	0.1%	99.7%
Disposable Diapers:							
HUGGIES	$186,118,300	667258994	313	$150,803,492	$2,517,733	37.3%	37.3%
PAMPERS	$177,271,897	640989147	313	$222,938,281	$10,812,085	35.6%	72.9%
PRIVATE LABEL	$75,562,803	385116429	313	NULL	NULL	15.2%	88.1%
LUVS	$50,635,624	221437197	313	$108,841,402	$4,881,835	10.2%	98.2%
DRYPERS	$5,343,677	24122528	313	NULL	NULL	1.1%	99.3%
FITTI	$2,531,552	12636144	313	NULL	NULL	0.5%	99.8%
SEVENTH GENERATION	$334,513	1008912	181	NULL	NULL	0.1%	99.9%
BENETTON	$300,838	1024892	178	NULL	NULL	0.1%	99.9%
LITTLE TIKES	$107,233	439944	71	NULL	NULL	0.0%	100.0%
BUMPIES	$77,220	401837	181	NULL	NULL	0.0%	100.0%

Table 5: Summary statistics for top ten brands in each category, for all weeks in database. "Dist", is number of weeks brand was present in at least one city. Volume is some metric of volume sold (not units), with market share and cumulative share being based on volume.

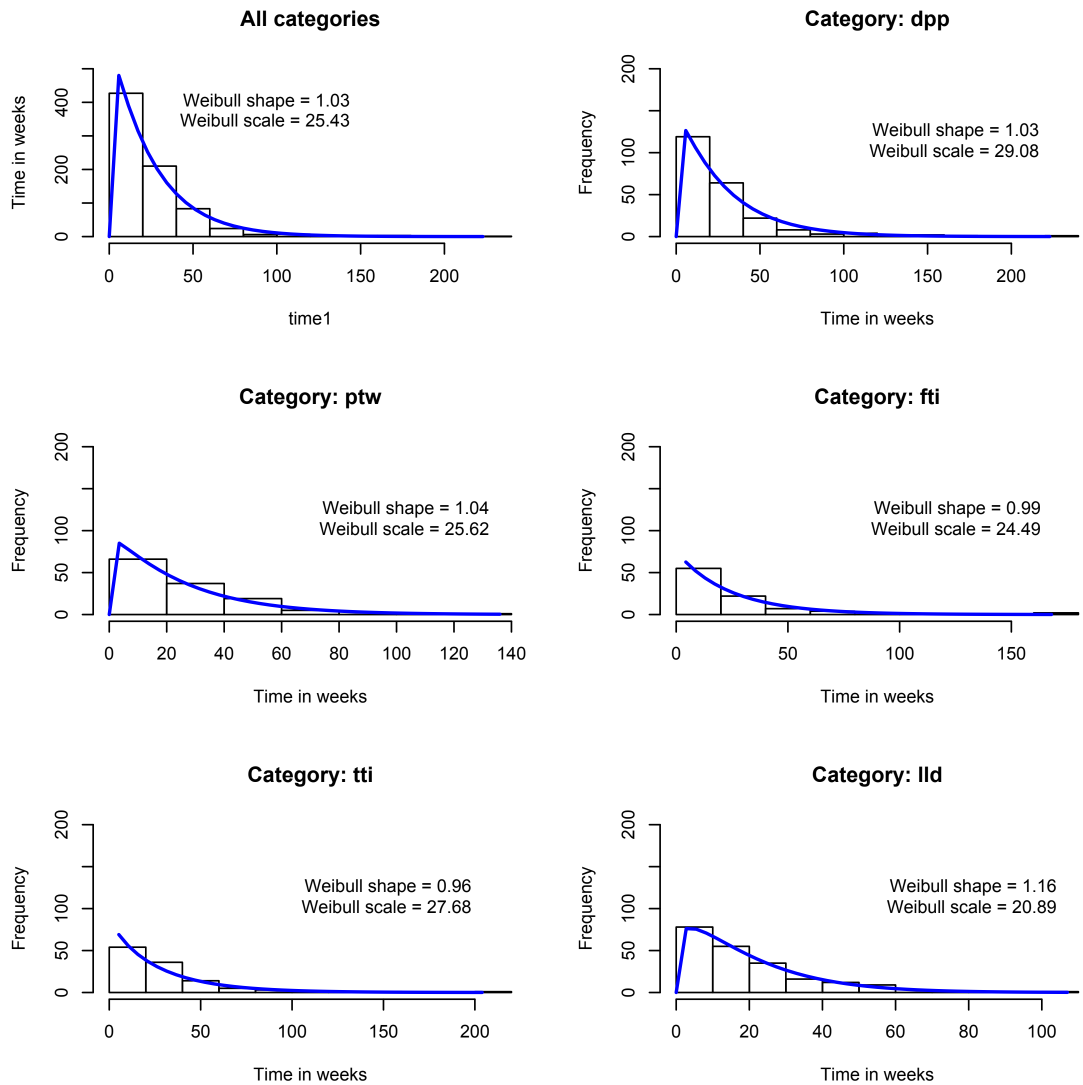

Figure 1: Density plots and overlay Weibull density values for duration of creatives used on TV. Each unit of analysis is a collapsed creative, being unique from other creatives.

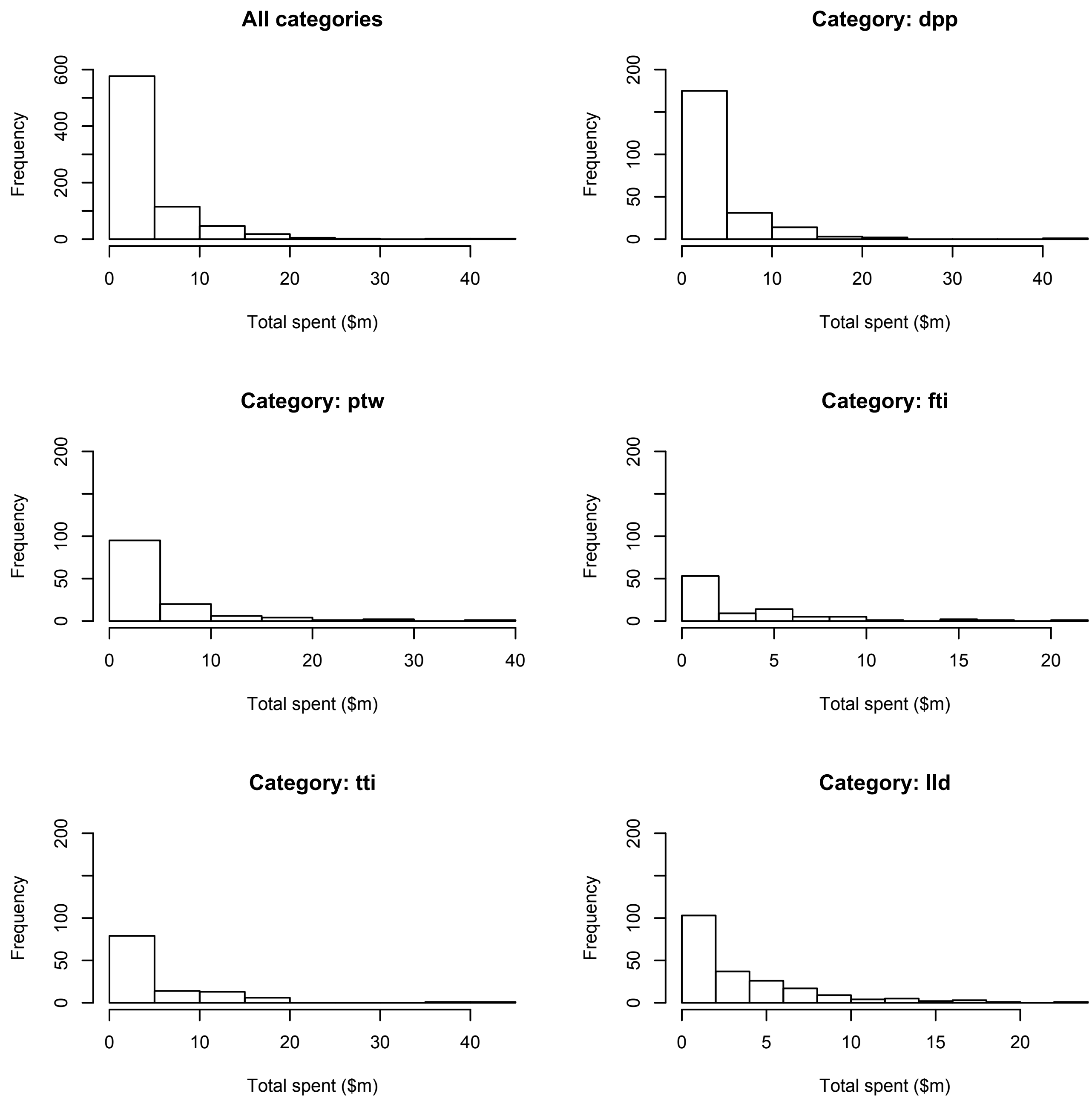

Figure 2: Total spent on new creatives over the lifetime of the creative being shown. The top left is across all categories, the other panels are by category.

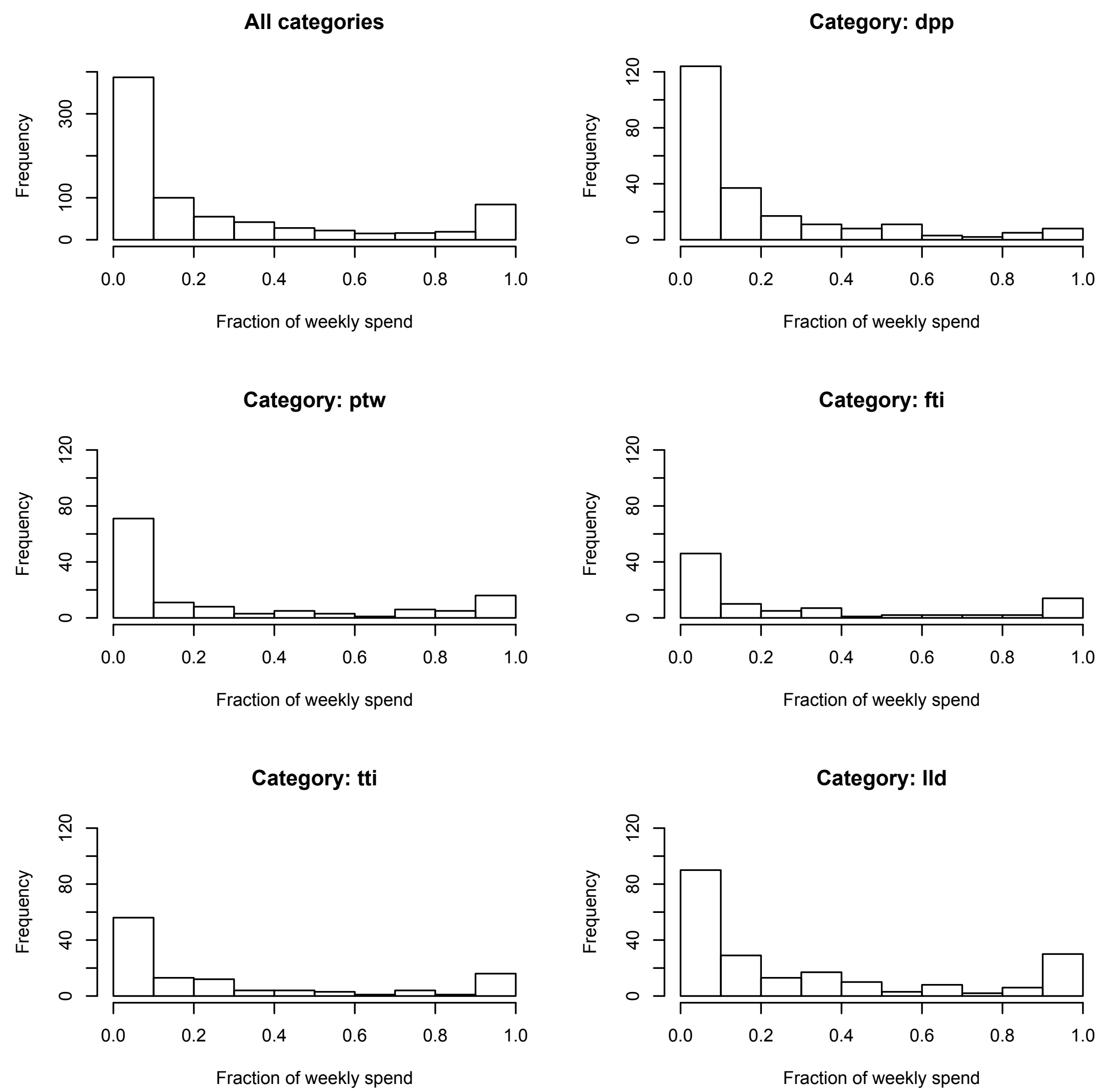

Figure 3: Histogram of fraction of the amount that the new creative for accounted for of the brand's total expenditure for the week. That is the total dollars spent on the new creative divided by the total spent by the brand on all advertising for that channel.

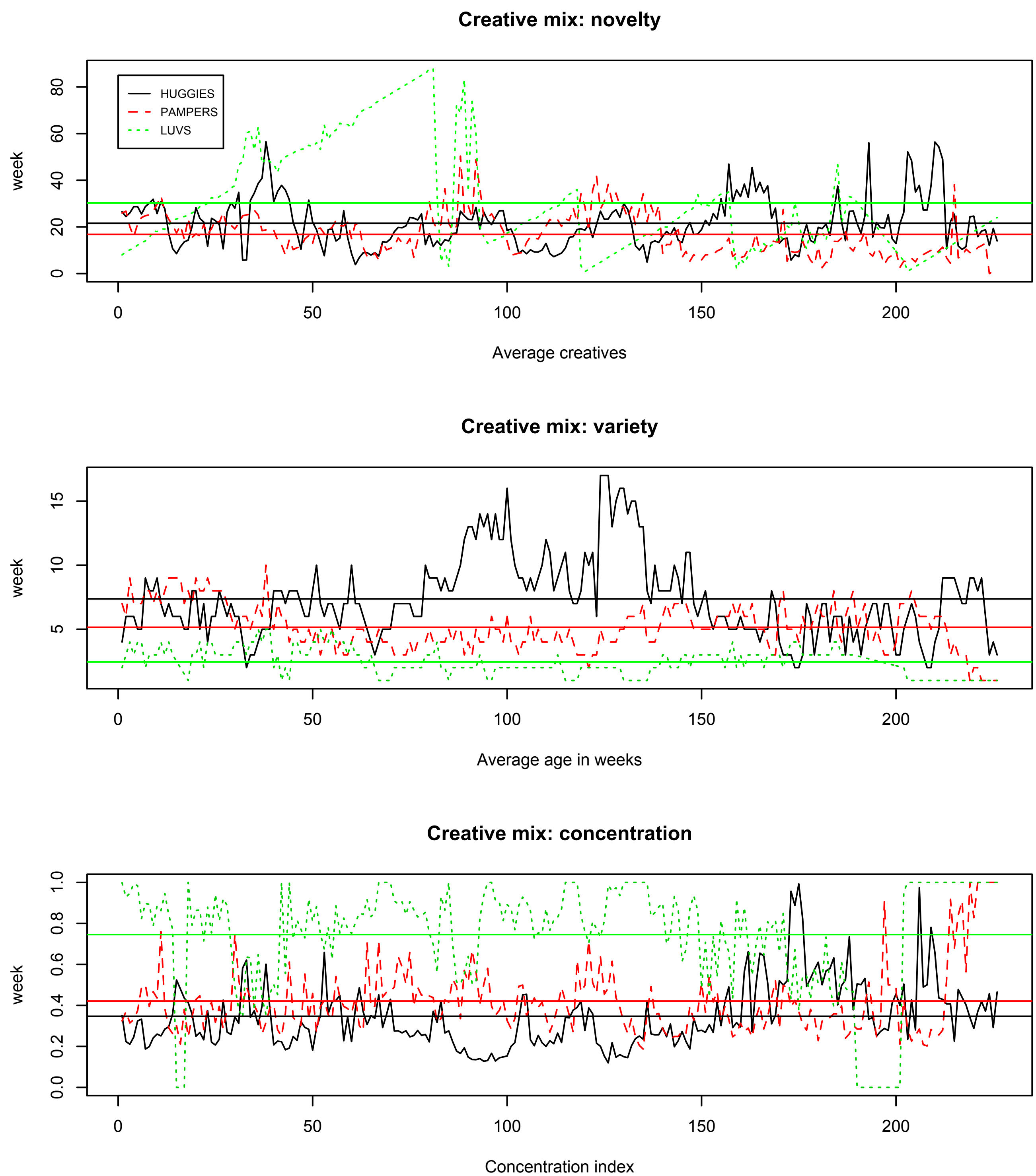

Figure 4: For the disposable diaper category, this displays the time series for the creative mix. The top panel is variety, the middle is novelty, and the bottom is concentration. The horizontal lines with corresponding color, are the means over time.