Timing and effects of new advertisement creatives

January 26, 2016

1 Introduction

Marketing organizations spend vast amounts of time and money each year on television advertising. The return on investment, being measurable as brand performance (e.g. see Ataman et al 200x), of this advertising expenditure hinges on how effective the campaigns are in a competitive environment (see e.g.). Over time, however, it has been observed numerous times that advertising can wear out, and this can be exacerbated by excess spending (e.g. NaikMantralSawyer1998 **?**, Bass et al 2008, Braun and Moe 2008). 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, Mantral 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. Braun and Moe (2010) were able to examine the introduction of new creatives on existing campaigns. Until recently, such models downplayed the role of competition. This was modeled in Naik, Prasad and Sethi (2009) in the context of brand awareness among a set of competing brands. In a mature category, and 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 of how it reduces 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, 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 competitive model of brand evolution, in the spirit of **?**, which includes dynamic advertising effects on sales. 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. 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 (compare with **?**. While the competing effects of brands has been considered in a more aggregate sense in e.g. **?**, we consider it explicitly among several of the major brands advertising. A study by **?** also considers competing brands' advertising, but model only the evolution of awareness, do not consider dynamics of campaign efficacy, and do not consider other marketing mix elements.

Very little empirical work exists on the dynamics of advertising response among several competing brands. We categorize the effects of advertising in this competitive environment on the basis of whether they act as defensive, versus offensive mechanisms (cites here). The central question of this research is then how implementing a creative changes these competitive dynamics. We present evidence that xxx. In our model, we control for 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. Our data 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 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. The combination of these two innovations is 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?

We also higlight that digital marketing may also benefit from this type of work, as marketers increasingly are turning to video based advertising, in favour of static appeals, as more effective ways to engage customers.

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. 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) and in response variables (any evidence here? Montgomery?). Further, we would be interested in capturing *covariation* among cities, again both in sales and in response to marketing instruments. The correlations among competing brands tend to be highest. For sales and market share, cross city variation is higher than variation over time.

Category	Nbrands	Revenue	price	fracdnp	fracfnp	fracdist	Nnewads	Mnewads
lld	8	486.93	5.68	0.13	0.16	0.80	283	56.60
tti	5	629.06	3.52	0.12	0.17	0.88	247	61.75
fti	3	213.77	1.54	0.10	0.23	0.90	124	62.00
dpp	4	337.24	11.72	0.04	0.26	0.84	302	100.67
ptw	6	509.99	2.76	0.11	0.13	0.84	165	33.00

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 represents a sample of stores and is aggregated across 42 cities. Price represents price per unit. For advertising, Nnewads is the number of new creatives in the overall category, Mnewads is average number of new creatives per brand, fracdnp/fnp is the fraction of UPCs available that were on display/feature, fracdist = average distribution (number of outlets).

TNS provided the data for national level advertising. While these sources were available across different media, such as cable, network, syndicated, and Spanish Language television, 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. (need a sense of how much it amounts to.)

A new creative is identified in our data as any change in creative description, as coded by TNS, that has

[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).

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.

Other than advertising, we collect or construct several other covariates that are associated with sales.

Pricing - 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 - retailer promotions include features and displays. Feature 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 - distribution is represented by the number of stores sampled, that have this brand. It is possible to use 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 number of UPCs within a 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.

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. This is an important innovation to this work, being the first to examine 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] We also used a text based metric for distance of descriptors to reduce the number of creatives to a smaller subset.

2. 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.

3. 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.

4. The change in advertising effectiveness is also governed explicitly by a change in creative appeal.

5. Other marketing effects - not all sales are due to television advertising. 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.

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).

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:

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

eqn:bqj1

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 $J_{bE} \subseteq J_b \subseteq J$)

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

eqn:bqj1

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 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:

$$Y_t = F_{11t}\Theta_{11t} + F_{12t}\Theta_{12} + v_{1t} \tag{3}$$

$$\Theta_{11t} = F_{2t}\Theta_{2t} + v_{2t} \tag{4}$$

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

[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:

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

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:

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

$$w_t \mid \Sigma, W \sim N(0, W, \Sigma) \tag{7}$$

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 ($\tilde{}$) 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.

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):

$$\begin{aligned} P(\bar{Y}, \Sigma \mid \mathcal{D}_0, \Psi) &= \prod_{t=1}^{T} P(\bar{Y}_t \mid \Sigma, D_{t-1}, \Psi) P(\Sigma \mid D_{t-1}, \Psi) \qquad (11) \\ &= \prod_{t=1}^{T} (2\pi)^{-\frac{NJ}{2}} |Q_t|^{-\frac{J}{2}} |\Sigma|^{-\frac{N}{2}} \exp\left[-\frac{1}{2} \operatorname{tr}\left(\left(\bar{Y}_t - f_t\right)' Q_t^{-1} \left(\bar{Y}_t - f_t\right) \Sigma^{-1} \right) \right] \\ &\quad \times IW(\nu_{t-1}, \Omega_{t-1}) \qquad (12) \end{aligned}$$

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

$$\begin{aligned} P(\bar{Y}|\cdot) &= \prod_{t=1}^{T} P(\bar{Y}_t | y_{1:t-1}, \cdot) \\ &= \mathcal{K} \left(\prod_{t=1}^{T} |Q_t|^{-\frac{J}{2}} \right) |\Omega_0 + \sum_{t=1}^{T} \left(\bar{Y}_t - f_t\right)' Q_t^{-1} \left(\bar{Y}_t - f_t\right) |^{-\frac{\nu_0 + TN}{2}} \qquad (13) \end{aligned}$$

eq:LL-T

where:

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

3.1 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

$$\Theta_{2t} = \begin{bmatrix} B_{1t} & B_{2t} & \dots & B_{Jt} \\ q_{11t} & q_{21t} & \cdots & q_{J1t} \\ q_{12t} & q_{22t} & \cdots & q_{J2t} \\ \vdots & & & \\ q_{1J_bt} & q_{2J_bt} & \cdots & q_{JJ_bt} \end{bmatrix} \qquad (14)$$

eqn:t2a

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:

$$F_{2t} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \vdots & & & \\ 1 & 0 & \dots & 0 \end{bmatrix} \tag{15}$$

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 (14) and (15) they combine to provide a $N \times J$ matrix which just depends on the intercept (B_{jt} in 2).

3.1.1 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 elasticity across cities):

$$\Theta_{2t} = \begin{bmatrix} B_{1t} & B_{2t} & \dots & B_{Jt} \\ q_{11t} & q_{21t} & \dots & q_{J1t} \\ q_{12t} & q_{22t} & \dots & q_{J2t} \\ \vdots & & & \\ q_{1J_bt} & q_{2J_bt} & \dots & q_{JJ_bt} \\ \theta^p_{11t} & \theta^p_{21t} & \dots & \theta^p_{J1t} \\ \theta^p_{12t} & \theta^p_{22t} & \dots & \theta^p_{J2t} \\ \vdots & & & \\ \theta^p_{1Jt} & \theta^p_{21t} & \dots & \theta^p_{JJt} \end{bmatrix} \tag{16}$$

[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 $P = J + J$

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

$$F_{2t} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ \vdots \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ \vdots \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{17}$$

3.2 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:

$$F_{11t} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & 0 & \ddots & \\ 0 & 0 & \dots & 1 \end{bmatrix} \tag{18}$$

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}:

$$F_{11t} = \begin{bmatrix} 1 & p_{11t} & p_{1jt} & \dots & p_{1Jt} \dots & 0 & & & & & \dots & 0 \\ \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & p_{n1t} & p_{njt} & \dots & p_{nJt} & \dots & 0 \\ \vdots & 0 & \ddots \end{bmatrix} \tag{19}$$

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

3.3 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. The function $\mathbb{I}(\cdot)$ is an indicator function that takes a value of 1 if the argument is greater than zero, and 0 otherwise.

$$\tilde{G}_t = \begin{bmatrix} (1-\delta) & \tilde{g}(A_{1t}) & \dots & \tilde{g}(A_{Jbt}) \\ 0 & 1 & \dots & 0 \\ \vdots & 0 & \ddots & \\ 0 & 0 & & 1 \end{bmatrix} \tag{20}$$

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:

$$H_t = \begin{bmatrix} 0 & \dots & 0 & \dots & 0 \\ \phi_{11}E_{1t} & \phi_{21}E_{1t} & \dots & & \phi_{J_b1}E_{1t} \\ \vdots & \ddots & & & \\ \phi_{1j}E_{jt} & \dots & \phi_{jj}E_{jt} & \dots & \phi_{Jj}E_{jt} \\ \vdots & & & & \\ 0 & \ddots & & & \\ \vdots & & & & \\ \phi_{1J_{bE}}E_{J_{bE}t} & \phi_{2J_{bE}}E_{J_{bE}t} & \dots & & \phi_{JJ_{bE}}E_{J_{bE}t} \end{bmatrix} \tag{21}$$

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.

3.4 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.

4 Additional to be done or considered:

- Include a hierarchical prior for the creatives, and allow the effect of the addition of creatives to be heterogenous within brands.

- Analytically derive the optimal number of creatives to run at any time, and look at how this depends on the variability in inital quality. Perhaps consider a choice of creative media agency? One with high variability, but low average quality, one with low variability? If it's a beta distribution we could also look at the media agencies that sometimes produce a hit, sometimes a complete flop, versus one that is more reliable "average" quality. A nice addition would be the allocation/policy for (possibly constrained) ad budgets, when the trade off involves selection of creatives, versus use of older/existing creatives and putting more money into reach/frequency objectives.

- Focus on initial qualities and multiple creatives per brand The accommodation of initial qualities could be a possible extension. Several possible outcomes could have some interesting implications. First, consider if the distribution across creatives is quite broad, i.e. there is wide heterogeneity in the quality of ads drawn (might want to discuss some of the creative agency literature). This would likely have a quite different outcome in terms of the decision to draw new ads (and the number of creatives being displayed) than if the heterogeneity is quite low. As a "hunch" we may expect the manager to want to try more creatives, sticking with the ones that seem to be high quality, but drawing new creatives if it is of low quality. We may also expect that the pattern of expenditure is quite different - with more pulsing behavior if the ad is high quality, but perhaps some initial pattern to 'learn' about the quality of the advertisement once introduced. Need to also consider how pre-testing of creatives may change these results. The problem with the estimation of this is in the size of the evolution matrix required which for just the advertising part (instead of a $J_b + 1$ by $J_b + 1$ matrix) requires a matrix which is of dimension equal to (with rather sloppy notation) $m + 1$ by $m + 1$ where m is the total number of creatives across brands.

Iterative estimation

To estimate the likelihood in Equation (13), 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.

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 2 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.

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

.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

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 2: Parameters to be estimated

tab:parameters

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:

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%
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%

Table 3: 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.

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	285
SPOKANE	79
ST. LOUIS	202
SYRACUSE	189
TOLEDO	137
TULSA,OK	96
WASHINGTON, DC	472

- 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.