Communicating Brands in Television Advertising

Norris I. Bruce¹
Maren Becker²
Werner Reinartz³

May 5, 2019

***Please do not cite, quote or distribute without permission***

¹ Associate Professor of Marketing, University of Texas at Dallas. Contact: The University of Texas at Dallas, Naveen Jindal School of Management, Office: SOM 13.310. Phone: 972-883-6293. Email: norris.bruce@utdallas.edu

² Assistant Professor, University of Cologne. Contact: University of Cologne, Faculty of Management, Economics, and Social Sciences, Department of Retailing and Customer Management, Albertus-Magnus-Platz 1, 50923 Cologne, Germany, Phone: +49 (221) 470-1325, Fax: +49 (221) 470-8677, e-mail: maren.becker@wiso.uni-koeln.de

³ Professor of Retailing and Customer Management, University of Cologne. Contact: University of Cologne, The Faculty of Management, Economics, and Social Sciences, Department of Retailing and Customer Management, Albertus-Magnus-Platz 1, 50923 Cologne, Germany, Phone: +49 (221) 470-8675, Fax: +49 (221) 470-5751, e-mail: werner.reinartz@wiso.uni-koeln.de
Communicating Brands in Television Advertising

ABSTRACT

Many studies have quantified the effects of TV ad spending or GRP (gross rating points) on brand sales. Yet this effect is likely moderated by the different types of brand-related messages or cues (e.g., logo, brand attributes) embedded in the ads and by the ways (e.g., explicitly or implicitly) these cues are conveyed to TV audiences. The authors thus measure 17 cues often used within ads to build brand awareness (or salience) and brand image and investigate their influence on ad effectiveness. Technically, the study builds a dynamic model to quantify the effects of advertising on sales; builds a robust and interpretable (i.e., non-parametric and sparse) factor model that integrates correlated, left-censored branding cues; and then models the effects of advertising as a function of the factors identified by these cues. An analysis of 177 campaigns aired by 62 brands finds that salience cues (e.g., logo) and explicit benefit and attribute messages moderate ad effectiveness and that explicit claims are more effective than implicit ones. The study can thus suggest ways brand and ad agency managers can improve the effects of creative ad content on sales.

Keywords: Ad Effectiveness, Ad Content, Brand Communication, Non-Parametric Factors, Dirichlet Process Mixtures, Sparse Tobit Factor Model, Sequential Monte Carlo Methods
INTRODUCTION

 Despite the emergence of online media, TV Advertising remains an important channel through which brands communicate with consumers (Batra and Keller 2016). Indeed, while Procter & Gamble Co. spent $7.1 billion and its main competitor, Unilever, spent almost $8.1 billion on ads in 2018 (Procter & Gamble Co. 2019; Unilever 2019), the majority of these efforts went toward TV advertising (Pellikan 2017). Not surprisingly, empirical research in this area has largely focused on quantifying the effects of ad spend or GRP (gross rating points) on brand sales (e.g., Assmus, Farley, and Lehmann 1984; Sethuraman, Tellis, and Briesch 2011). Yet the effectiveness of brand advertising is likely moderated by the different types of brand-related messages or cues (e.g., logo, brand attributes) embedded in the ads, and by the ways (e.g., explicitly or implicitly) those cues are conveyed to TV audiences. This paper thus develops a set of flexible empirical models to quantify how brand messages or cues embedded in TV advertising influence the effects of ad spending on brand sales. After all, brand and agency managers need to know not only if ad spending works (Tellis 2004) but also how brand messages (e.g., an agency’s creative input) contribute to that outcome.¹

 Of course, the goal of advertising is commonly to build brand awareness and/or to present favorable brand images, which has the long-term potential to create brand equity in the form of a loyal base of consumers (Keller 1993). Yet, efforts to increase (say) awareness, such as making a brand overly salient in a TV ad (e.g., Rossiter and Percy 1997), may annoy consumers;

¹Gordon Euchler, planning manager at BBDO, states that during the planning process, agency and brand managers would inevitably discuss how to effectively integrate brand messages into an ad. However, he noted that while brand managers are mostly concerned about an ad’s effect on sales (Bruce, Peters, and Naik 2012), the agency is primarily concerned with the persuasiveness and likeability of the ad story; still, an ad being liked due to a creative story should ultimately influence sales; otherwise, they would eventually lose the account.
they may then engage in counterarguments or ad avoidance (Teixeira, Wedel, and Pieters 2010), which could in turn diminish the effects of the advertising. Furthermore, to create brand awareness, firms can deploy numerous cues within TV ads (e.g., using sound and images), such as frequent mentions of the brand name or integrating the logo and/or product. Thus, if branding cue salience enhances TV ad effectiveness, it remains useful for brand or agency managers to quantify which cues determine this outcome. Brand advertising may also stress favorable attributes and benefits (i.e., brand associations) to strengthen brand image. Even so, for low-involvement brands, it is unclear whether consumers are sufficiently motivated to process such information (MacInnis, Rao, and Weiss 2002). Because TV advertising permits immense creative flexibility (due to the many possible combinations of sight, sound, color, motion, and drama), firms can employ a variety of association cues within their campaigns (e.g., product-related vs. non-product-related; functional, experiential, symbolic benefits). Yet, they need to choose these cues judiciously, given known concerns about commercial clutter, consumers’ limited cognitive capacity, and the short duration of TV ads.

Guided by the above issues, we investigate the following questions for a set of fast-moving consumer goods (FMCG) brands:

- Will efforts to make a brand more salient in TV ads increase ad effectiveness? Which salience cues (e.g., frequency of brand name mentions; duration of time the logo is displayed) are most effective?
- Similarly, does featuring brand association cues enhance ad effectiveness? If so, should brands focus on attributes or benefits (or both)? What branding cues (e.g., product-related, non-product-related, explicit, or implicit) are most effective? Should they be explicitly or implicitly conveyed?
• How can brand managers and ad agencies use the results of this analysis to improve brand communication in TV advertising and, in turn, improve sales?

To address these questions, we first build a dynamic ad response model (Bass et al. 2007) that quantifies the effects of TV ads on brand sales. We then build a flexible (robust) factor model to obtain a set of latent factors and interpretable parameters (loadings) from multiple (noisy, correlated) branding cues mined from TV ads seen by viewers. The latent factors moderate the effects of TV ads on sales, and the loading parameters measure the relative importance of each branding cue to 1) the identification and interpretation of each factor, and 2) ultimately, to the success of advertising.

Naturally, depending on their communication goals (image or awareness building), brands may stress similar cues very differently across ad campaigns, and may employ a given cue (e.g., price) in some campaigns but not others. Hence, mining the content of TV ads for this study results in cue data that are non-Gaussian (e.g., skewed, multi-modal) with a large proportion of zeros (left-censored). This means we cannot reliably measure the effects of branding cues in TV ads using traditional factor methods, which often assume latent factors are Gaussian, and that restriction could contaminate parameters in both factor and ad-sales models (Piatek and Papaspiliopoulos 2018). Thus, we innovate to build a flexible or robust (i.e., Bayesian non-parametric (NP)) factor model for the branding cues.² A first step toward NP modeling of the factors is to use the widely applied Dirichlet Process Mixture (DPM) framework (see Dey et al. 1998; Escobar and West 1995, 1998; Bruce 2019). In this case, we assume latent factors emerge from a distribution that is itself uncertain (i.e., its shape is flexible) and that distribution emerges from a DP (Carvalho et al. 2008). This gives a flexible specification that

² We thank an anonymous reviewer for encouraging this innovation.
will adapt to the radically non-Gaussian structure (later) evident in the mined branding cues. Finally, we employ Tobit models to help resolve the zero observations (left-censoring) problem in many cues; similar methods in marketing arise from the analysis of purchase volumes across multiple categories. Treating these zeros as missing is again incorrect, for the censoring mechanism that produces zeros contains information about the factors (Kamakura and Wedel 2001).

Furthermore, to obtain loadings that facilitate a “cleaner” interpretation of the relative importance of different cues, we adopt a Bayesian sparsity (i.e., regularization) method (e.g., West 2003 and Carvalho et al. 2008) for the estimation of the factor loadings. This is analogous to Varimax rotation (Kaiser, 1958), which yields loadings that are either very small (near zero) or large (Rockova and George 2015). Naturally, the Bayesian approach works via priors (e.g., spike-slab distributions) that induce sparsity; these allow arbitrary patterns of zeros among the factor loadings parameters, such that the cues (data) inform the actual patterns. Moreover, sparsity has other advantages that support robust inference. If the factor loadings are sufficiently sparse (as they turn out to be in this study), one can uniquely identify latent factors (Kaufmann and Schumacher 2019) without designating specific (“founder”) cues for interpretation and/or identification (Carvalho et al. 2008). The latter strategy, popular in economics and marketing (e.g., Cunha, Heckman and Schennach 2010; Bruce, Peters, and Naik 2012\(^4\)), anchors latent factors to real observations; but the founder cues then have considerable weight in determining the factors and so could bias our estimates. Our sparse Bayesian methods thus provides a robust way to jointly infer the number of factors while estimating the ad-sales model, which avoids the

---

3 We thank the AE for this helpful suggestion.
4 Bruce, Peters, and Naik 2012 used survey responses (on 7-point scale) to “I like advertising of this brand very much” to identify and name an “affect” factor.
misspecification problem that can arise if one were to choose the number of factors or designate restrictions \textit{a priori}.

We calibrate the ad-sales model with weekly retail scanner panel and media data for 62 brands and 177 ad campaigns; these span six FMCG categories sold in Germany over approximately four years. The data provide weekly brand sales, TV ad spending, and information on various control variables, such as price, in-store promotions, and expenses for other advertising activities (i.e., internet, billboard, and print); thus, we account for heterogeneity and potential sources of endogeneity in the sample. In contrast, to quantify the moderating effects of multiple brand messages, we had to obtain branding cues directly from TV ads seen by viewers; so, we employed a set of trained experts to code all 177 advertising campaigns in terms of 17 branding cues, which become inputs into our factor model. We estimate both models (again, factor and ad-sales) jointly using a Bayesian approach to the Kalman filter (Bass et al. 2007) and inference for sparse loadings and DP latent factors (Carvalho et al. 2008). Notably, traditional inference would involve a two-pass procedure: in the first pass, we estimate the loadings and factors; in the second, we estimate the ad-sales model, using estimates of factors as exogenous variables. This later pass would suffer an errors-in-variables problem, because it ignores the uncertainty of the factor estimates; this can potentially lead to incorrect inference. Bayesian estimation obviates this problem (Geweke and Zhou 1996).

In summary, this study quantifies the effects of cues in TV ads on brand sales, using the cues often employed to build brand awareness and image associations (e.g., Keller 1993; Rossiter and Percy 1997). To do this, we introduce methods for non-parametric, sparse, Tobit factor models. Sparsity gives us managerially \textit{interpretable} (“cleaner”) loadings similar to those obtained from \textit{varimax} (Kaiser 1958) post-processing. Sparse inference also facilitates model
identification without the need to designate specific (“founder”) cues for naming and factor identification (Bruce, Peters, and Naik 2012). Our factor method is thus a “hybrid” approach that combines the parsimony of confirmatory analysis while retaining some of the flexibility of exploratory analysis (Lu, Chow, and Loken 2016). That means, moreover, that we are able to do factor exploration jointly with the estimation of the effects of advertising on sales, without using multi-pass procedures that ignore the role of factor uncertainty on these effects. Our cue data also present censored and non-normal measures, and so we use a Tobit specification for some cues and assume factors emerge from DPMs (e.g., Bruce 2019) to mitigate these concerns. Results show, for example, that salience and some association cues have positive effects on TV advertising but that brand communication is most effective when managers combine salience cues (primarily brand logos) with explicit benefit and attribute cues, and that explicit claims are more effective than implicit ones. Notably, advertisers often combine explicit product attributes with benefit claims, seemingly giving consumers reasons to buy the brand (Aaker 1991; Keller 2007). Similar studies have largely focused on using lab experiments to determine the effects of selected cues (e.g., frequency or timing of brand name mentions) on mindset measures, such as recall, attitude, or purchase intentions (Baker, Honea, and Russel 2004; Romaniuk 2009); to the best of our knowledge, none provides tactics that can moderate the effects of brand messages within ads (e.g., Draganska, Hartmann, and Stanglein 2014). Our study can help brand managers and ad agencies track, and potentially improve, their brand communications; and our methods should be useful to marketing scholars who wish to employ (robust) NP Bayesian and/or sparse factor analyses of high dimensional (big) data.

LITERATURE REVIEW

Impact of Advertising Content
We note two streams of research on the role of ad content: one that quantifies its effects on the 
ad-sales relationship in real markets, as done in this study, and another that quantifies its effects 
on mind-sharing metrics (e.g., recall, liking) in lab experiments. Using the first approach, 
Eastlack and Rao (1989) (and later Lodish, Abraham, and Kalmenson 1995) find that increasing 
the level of spending does not necessarily enhance ad effectiveness, but changes in ad content 
can affect sales. More recently, Liaukonyte, Teixeira, and Wilbur (2015) find that ad content also 
influences consumers’ online shopping behavior. Other studies in this stream report the 
moderating roles of broad types of ad cues (appeals) in different product-market settings. For 
example, in established product categories, creative and emotional cues are more effective than 
informational cues (e.g., Bass et al. 2007; Becker, Wiegand, and Reinartz 2019; Chandy et al. 
2001; MacInnis, Rao, and Weiss 2002). Studies that use experiments, which have the inherent 
benefits of control and internal validity, have contributed numerous findings (e.g., see summaries 
in Chattopadhyay and Basu 1990; Loewenstein, Raghunathan, and Heath 2011; Morales, Wu, 
and Fitzsimons 2012). Nonetheless, we know that lab experiments may not account for common 
market conditions, e.g., i) they may not account for competition or other market forces; ii) it is 
often infeasible to test multiple content cues within a single lab study, yet TV ads often encode 
multiple branding cues, given the richness of the medium; and iii) lab studies force respondents 
to process ads actively, whereas consumers in real markets often do so passively.

Thus, while marketing texts (Rossiter and Percy 1997) provide useful practitioner 
guidelines for the use of specific branding cues (termed “creatives”), to the best of our 
knowledge, no published study has quantified how brand messages embedded in TV ads 
influence the effects of ad spending on brand sales; however, we do note two studies that are 
broadly related. The first study, Teixeira, Wedel, and Pieters (2010), which uses eye tracking to
analyze the effects of branding cues on ad avoidance, finds that prominently featuring the brand increases avoidance but that pulsing (i.e., showing the brand frequently for short durations) reduces it. Their research is one of the few attempts to study multiple branding cues in advertising. They however focus exclusively on branding cues that generally contribute to awareness and consider their impacts on ad avoidance; we in contrast focus on the sales effect of TV ads, and furthermore, on the effects of attribute and benefit messages. The second study, Bruce, Peters, and Naik (2012), employs mind share measures to identify the effects of latent factors, cognition, affect, and experience (i.e., the intermediate effects of ads) on the sale of a single FMCG brand. The novelty of our study is that we focus on the brand content of real TV ads, so we can assess how managers should design ads with different branding cues in order to increase returns on ad spending.

**CONCEPTUAL MODEL -- THE EFFECTS OF BRANDING CUES IN TV ADS**

We now propose a conceptual model for this study. That is, we broadly define the relevant cues and the limitations of exhibiting them in TV ads, and then outline a conceptual model of how these cues moderate the effect of ad spending on sales.

First, marketing scholars and practitioners largely agree that successful ads should build awareness and give consumers reasons to buy the brand (Aaker 1991; Keller 2007; Teixeira, Wedel, and Pieters 2010). The cues that help promote awareness (i.e., salience cues) include mentioning the brand name, displaying the logo, or showing the product in the ad. The idea is that the more frequently consumers hear, see, or think about the brand, the more prominent it becomes in their memory (Elliot and Percy 2007), which in turn could lead to sales. Yet, as mentioned earlier, it is unclear whether featuring cues prominently improves ad effectiveness (e.g., Rossiter and Bellman 2005; Teixeira, Wedel, and Pieters 2010); it could instead annoy
consumers and prompt them to generate counterarguments or even avoid the advertisement (Teixeira, Wedel, and Pieters 2010). Furthermore, exhibiting branding cues in ads for familiar brands may be less important because consumers can effortlessly activate existing knowledge of these brands (Elliot and Percy 2007).

On the other hand, brands often embed association cues in their TV ads to help persuade consumers to buy; these cues can contain both attribute and benefit statements (Elliott and Percy 2007; Stewart and Furse 1986). Attribute cues define the objective features of a brand. These can be product-related features that relate directly to the product’s performance, such as ingredients (e.g., 100% organic, fresh oranges); and non-product-related features that do not directly affect performance but relate to the general product experience, such as price or packaging. Benefit cues however are “the personal values consumers attach to the product attributes” (Keller 1993, p. 4). They can be functional, experiential, or symbolic. Functional benefits stress the inherent advantages of product usage and address the problem-solving needs of consumers (e.g., cleans, removes dandruff) (Park, Jaworski, and MacInnis 1986); experiential benefits describe the sensory pleasure that consumers can derive from product consumption (e.g., fragrance or taste) and how it feels to use the product (Keller 1993); and symbolic benefits relate to the extrinsic advantages of using a product, such as prestige, personal expression, or social approval (e.g., shiny hair, attractiveness to women, enhanced self-esteem). Lastly, brands can emphasize attributes and benefits explicitly or implicitly; explicit claims are more direct, whereas implicit claims are more subtle and allow the customer more room for interpretation (Kardes 1988, Sawyer and Howard 1991). Thus, marketers may prefer to communicate some

---

5 We assume that brand association cues within advertisements have positive connotations.
branding cues (e.g., product-related and functional cues) explicitly and others (e.g., symbolic and experiential cues) implicitly.

It follows, then, that with these many options, marketers may need to know if cues are effective and how cues work collectively to influence advertising. For example, prominently featuring the brand name, logo, or product might be sufficient to enhance ad effectiveness (MacInnis, Rao, and Weiss 2002) and in turn sales, especially for low-involvement brands, because consumers of such brands are less apt to process product information. In addition, some attributes and benefits are likely related, so it might be redundant to integrate both types of cues into an ad (Wu, Day, and MacKay 1988). It is also unclear whether the different association cues should be presented explicitly, implicitly, or both. Studies suggest that implicit cues can lead to more positive ad response (e.g., Kardes 1988) when consumers are motivated and able to process ad messages. Yet in our context (TV ads for low-involvement FMCG brands) it is unlikely that consumer will be motivated or have the cognitive resources to attend to ad messages due to the presence of ad clutter\(^6\) (Sawyer and Howard 1991). Finally, TV advertising permits immense creative flexibility; brands can thus employ a variety of benefit and attribute cues. Yet they may need to discriminate among many options, given concerns about commercial clutter and the short duration of TV ads (usually around 20 seconds).

To address these and the earlier data issues, we propose an empirical model, illustrated conceptually in Figure 1. That is, we infer a number of \(k\) latent factors from censored and non-Gaussian branding cues \((y_{1..17})\) mined from the TV advertising of multiple FMCG brands. We then quantify the influence of these \(k\) factors on \(\beta\), the effect of advertising input \(a\) on brand

\(^6\) Gordon Euchler, planning manager of BBDO Germany, argues that consumers are more skeptical about implicit claims. According to him, they tend to suspect that the reason brands don't explicitly state something is because it's just not true.
sales \((R)\). With this model, brands can identify the relative effects of different cues in TV advertising via a set of interpretable factor loading parameters \(\lambda\). We now describe the empirical model and later its estimation.

---Insert figure 1 here---

**MODEL DEVELOPMENT**

Recall that we aim to accomplish two broad tasks: first, model the dynamic relationship between brand advertising and sales (e.g., Bass et al. 2007); second, innovate to develop a robust (flexible) factor model for left-censored and non-Gaussian cues \((y_1-y_{17})\) that infers a number of \((k)\) latent factors. Formally, we adopt a Tobit factor model (e.g., Wedel and Kamakura 2001) for the cues where latent factor distributions are themselves uncertain but emerge from Dirichlet Process Mixtures (DPMs); this flexible, non-parametric density will automatically adapt to non-normal cue data (e.g., See Carvalho et al. 2008; Bruce 2019). Moreover, because our study requires managerially interpretable loading patterns, we use Bayesian methods (e.g., West 2003; Carvalho et al. 2008) to estimate sparse factor loadings (equivalent to frequentist Varimax post-processing), which also helps identify the relevant \((k)\) factors (e.g., Lu, Chow, and Loken 2016; Kaufmann and Schumacher 2019). Naturally, the model should also control for heterogeneity and potential endogeneity.

**Dynamic Advertising Response Model**

Equations 1 (sales) and 2 (goodwill) define our ad response model for a brand \(i\). They incorporate the current effects \(\beta_{ij}\) of multiple \((j = 1, 2, \ldots, J_i)\) ad campaigns\(^7\) and the long-term or carryover \(\delta_i\) from past advertising on the brand’s goodwill \(G_i\), which in turn affects brands sales \(R_i\). The variables in \(z\) (equation 1) capture the effects of other forces on sales, such as the

---

\(^7\) Here, a campaign represents a set of commercials (i.e., executions) with a common theme.
focal brand’s price, promotion, and other advertising activities, as well as the price and ad spend of competitor brands. Several variables, including own price, could be endogenous; we return to this issue subsequently. To specify diminishing returns to advertising, we use the log transformation \( g(a_i) = \log(1 + a_i) \) (See e.g., Bass et al. 2007). Here, the goodwill intercept parameter \( G_i \) represents the mean level of initial brand goodwill in the absence of advertising (Naik and Raman 2003; Ataman, van Heerde, and Mela 2010). Lastly, the measures \( w_{0it} \) and \( w_{it} \) denote normal specification errors in the sales and goodwill equations, respectively.

\[
R_{it} = G_{it} + z_{it} \eta_i + w_{0it}
\]

\[
G_e = G_r + \delta G_{r-1} + \sum_{j=1}^{J_r} \beta_j g(a_g) + w_e
\]

where \( i = 1, \ldots, N \) brands; \( j = 1, \ldots, J_i \) campaigns \( t = 1, \ldots, T \) weeks; \( w_{it} \sim N(0, \sigma_i^2) \); and \( w_{0it} \sim N(0, \sigma_{0it}^2) \).

Next, to specify how multiple branding cues moderate TV ad effectiveness, we model the effectiveness of brand \( i \)'s campaign \( j, \beta_{ij} \), as a linear combination of \( k \) relevant latent cue factors \( f_{jh} (h = 1, \ldots, k) \), a set of control variables \( X_{ij} \), and a brand specific intercept \( \alpha_{0i} \). While the intercept measures unobserved heterogeneity, the control variables, \textit{emotional appeal} and \textit{line extensions}, measure the emotional content of the ad and whether the ad promotes a new line extension (see details in Table 4). These controls, in particular \textit{emotional appeal and line extensions}, may influence ad effectiveness (e.g., Chandy et al. 2001; MacInnis, Rao, and Weiss 2002). Thus, we specify the effectiveness of an ad campaign, \( \beta_{ij} \) as follows:

\[
\beta_{ij} = \alpha_{0i} + \sum_{h=1}^{k} \lambda_{0ih} f_{hij} + \mu X_{ij} + \nu_{0ij}, \quad \alpha_{0i} \sim N(\alpha_0, \sigma_0^2)
\]
where $\lambda_{0h}$ and $\mu$ are parameters for the latent factors and control variables and $\nu_{0ij}$ is a normally distributed random noise, $\nu_{0ij} \sim \mathcal{N}(0, \sigma^2)$. Note that if $\lambda_{0h}$ is significant then the cues loaded on factor $h$ significantly influence ad effectiveness. It is also important to anticipate the robustness (flexibility) of this specification (3). It can incorporate the effects of a large number of correlated, noisy cues via a smaller set of common factors. The number of relevant factors, $k$, is also unknown, $a$ priori, but determined during model estimation. Lastly, although traditional factor models in economics and marketing assume factors $f_{hij}$ emerge from normal distributions, we later relax this restriction.

A Robust Factor Model for Branding Cues

We now specify a robust factor model for censored, non-normal branding cues. Again, the factor approach provides a parsimonious way to incorporate many correlated cues into the ad-sales model, equations 1-3. Nonetheless, standard factor methods seem inadequate for this study. For example, to make inference feasible and obtain interpretable results, analysts commonly use identifying assumptions, often parametric (“hard”) restrictions (e.g., loading parameters) and strict distributional assumptions (e.g., normal factors); but here, these can bias our findings, for the reasons given earlier. Thus, we propose a Bayesian method that induces sparsity through the use of priors (or “soft” restrictions, via spike and slab priors; defined later) that allow arbitrary patterns of zeros in the factor loadings matrix, such that the branding cues (i.e., the data) inform the patterns (e.g., Lu, Chow, and Loken 2016; Kaufmann and Schumacher 2019). Similarly, for identification, all published marketing studies thus far assume factors arise from normal distributions. Yet our branding cues are non-normal, and so we assume factor densities are non-parametric (NP) (e.g., Dey, Müller, and Sinha 1998; Escobar and West 1995;
Lastly, we adopt a Tobit specification to mitigate the left-censoring process in our branding cues (e.g., Wedel and Kamakura 2001). To simplify our exposition, we begin with a typical factor specification and then incorporate censoring, sparsity, and NP density assumptions. That is, for each brand \( i \) and campaign \( j \), the vectors \( y_{ij} \) of 17 branding cues and \( k \) latent factors \( f_{ij} \) have a linear relationship through a matrix \( A \) of factor loadings and a brand intercept vector, or unobservables that can influence cue decisions, \( \zeta_{ij} \):

\[
y_{ij} = \zeta_{ij} + Af_{ij} + v_{ij}, \quad \text{where}
\]

\[
f_{ij} \sim N(0, \Sigma_{f_i}), \quad v_{ij} \sim N(0, \Sigma), \quad \Sigma = \text{diag}(\sigma_f^2, \ldots, \sigma_f^2), \]

\[
A = \begin{bmatrix}
\lambda_{1,1} & \cdots & \lambda_{1,k} \\
\vdots & \ddots & \vdots \\
\lambda_{17,1} & \cdots & \lambda_{17,k}
\end{bmatrix}, \quad 17 \times k \text{ loadings matrix.}
\]

As noted earlier, several attribute and benefit cues include a large proportion of zeros; treating these as missing is inappropriate because the censoring mechanism producing the zero may contain information about the factors. For cues \( c=1,\ldots,17 \) that include zeros, we therefore adopt a type I Tobit specification: we augment equation (4) with the latent (cue propensity) measure \( y_{ij}^* \), such that \( y_{ij}^* = y_{ij} \) if \( y_{ij} > 0 \); otherwise \( y_{ij}^* = \zeta_{ij} + Af_{ij} + v_{ij} < 0 \), where \( y_{ij} \) is observed only if it is larger than zero (See e.g., Wedel and Kamakura 2001 for a Tobit Factor model). For estimation, we take the log of the positive cue values (Cameron and Trivedi 2005, pp. 531), to reflect findings that cue effects diminish the more they are employed in an ad (e.g., Teixeira, Wedel, and Pieters 2010). For simplicity, we further assume the error components \( v_{ij} \) are independent, normal densities. Independent errors are standard in factor analyses; they imply that the factor term \( (Af_{ij}) \) is the source of correlation among the cues. The other default
option in the literature is that latent factors $f_i$ are normal, but we later assume factors emerge from a non-parametric density. Again, $k$, the number of relevant factors, is unknown, \textit{a priori}.

Yet the factor model outlined in Eq. (4) is unidentified. In particular, for any non-singular $k \times k$ matrix ($B$), the transformation $Af_i = (AB)(B^{-1}f_i)$ leaves the distribution of the cues $y_i$ unchanged. To identify equation (4), we need $k^2$ restrictions. Analysts often address this “rotational” problem by assuming the factors ($f_i$) are uncorrelated (and normal) with unit variance, $\Sigma_f = I$, which gives us $k(k + 1)/2$ restrictions; but that means factor identification requires another $k^2 - k(k + 1)/2 = k(k - 1)/2$. One popular way to obtain the remaining $k(k - 1)/2$ restrictions is to set the upper-diagonal elements of the factor loadings matrix to zero (e.g., West 2003). Alternatively, one can restrict variables to load on to specific columns of the factor loading matrix (e.g., Carvalho et al. 2008). The latter approach, popular in the econometrics and marketing literatures, “anchors” the latent factors in real (“founder”) measurements, thus facilitating both \textit{identification} and \textit{naming} of the factors (e.g., Cunha and Heckman 2008; Cunha, Heckman, and Schennach 2010; Bruce, Peters and Naik 2012). But these “founder” variables in turn have considerable influence in determining the factors and so limit the robustness of one’s findings, especially when one has no prior information to help select “founder” measures. In this study, we avoid these arbitrary restrictions, for in our cue data, the degree of sparsity is sufficiently high to obtain the additional $k(k - 1)/2$. (See, e.g., Kaufmann and Schumacher (2019) for identification strategies when sparsity is insufficient).

\textit{Sparse Factor Loadings:} In Bayesian factor analyses, a sparse loading matrix contains (some or many) zeros in its columns and rows. In fact, entire rows and columns of zero loadings indicate irrelevant variables and factors, respectively; that means we can infer both the relevant $k$
factors and the branding cues \( y \) (and potentially obtain identifying restrictions) while estimating a sparse factor model. Conversely, non-zero loadings in columns indicate the variables on which the factors load, potentially assigning interpretations to the factors. This is the main advantage of sparsity in this study. Our method hence combines features of both exploratory and confirmatory factor analyses. Naturally, there are other ways to implement sparsity or near-sparsity in the loading matrix. In frequentist work, one might use a sparse eigenvalue decomposition (Zhu, Yu, and Zhu 2010) or post-hoc varimax rotation (Kaiser 1958).

Here, we adopt the Bayesian approach of West 2003 and Carvalho et al. 2008, who applied sparse models to gene expression studies. The method induces sparsity in the factor loading matrix \( A \) via \textit{two-level} (spike and slab) \textit{mixture priors}. As a result, we can obtain interpretable loadings during model estimation, rather than \textit{post-hoc}, as done with Varimax rotation. Each level of the hierarchy is a mixture of a Dirac delta function with infinite mass at zero (i.e., the \textit{spike}, the prior belief a parameter is zero) and i) a normal for the factor loadings \( \lambda_{ch} \) (\textit{cues}, \textit{c}; \textit{factors}, \textit{h}=1,...,\textit{k}) or, ii) beta distribution for a variable specific probability \( \pi_{ch} \) of a non-zero loading \( \lambda_{ch} \neq 0 \) (i.e., the \textit{slab}, the belief a parameter deviates from zero). These priors help introduce patterns of zeros among the factor loading elements \( \lambda_{ch} \), revealing the relative importance of the branding cues. Posterior updates then yield either zero loadings or loadings shifted away from zero.

More formally, the first level of the hierarchy is the point mass mixture prior for the elements of the loadings matrix:

\[
\lambda_{ch} = (1 - \pi_{ch}) \delta_{\lambda_{ch}}(0) + \pi_{ch} N(\lambda_{ch} | 0, \tau)
\]
where the spike $\delta_0(\lambda_{ch})$ is the Dirac delta function at zero. The non-zero loading elements $\lambda_{ch}$ for branding cue $c$ and factor $h$ arise from a normal prior with mean 0 and variance $\tau_h$ (the slab). At the second level, the variable specific probabilities $\pi_{ch}$ have similar mixture priors,

$$
\pi_{ch} = (1 - \rho_h)\delta_0(\pi_{ch}) + \rho_h Be(\pi_{ch} | b_h m_h, b_h (1-m_h))
$$

where $Be(\pi_{ch} | b_h m_h, b_h (1-m_h))$ is a beta distribution with mean $m_h$ and precision $b_h > 0$. The factor probabilities $\rho_h$ have a prior that heavily favors very small values, such as $Be(s | sr, s(1-r))$, where $s > 0$ is large and $r$ is a very small prior probability of non-zero values. The use of the above variable specific loading probabilities $\pi_{ch}$ (rather than common ones, $\pi_{ch} = \pi_h$) allows the model to obtain “cleaner loadings”; namely, to much more effectively (and automatically) detect non-zero loadings (See West 2003; Carvalho et al. 2008) and to induce shrinkage towards zero for elements of the factor loading matrix, $A$. This automatic adaptation is a key advantage of the Bayesian approach, which obviates the need for ad-hoc or post-hoc (e.g., Varimax) methods. The posteriors, $\pi_{ch} = \Pr(\lambda_{ch} \neq 0 | y)$, which reflect cue-factor associations, help us rank and select cues associated with factor $h$. We report loadings $\lambda_{ch}$ with large posterior probabilities,

$$
\pi_{ch} = \Pr(\lambda_{ch} \neq 0 | y) > .99 \text{ (See Web Appendix W1 for posterior computations).}
$$

**Non-Gaussian (Non-parametric) Factors:** Within marketing, all published papers so far assume factors to be Gaussian. While this assumption is convenient, it has insufficient empirical justification in this paper. As noted, our branding cues are non-Gaussian, and so traditional factor analysis (with strict normal factors) would lead to poor model fit and likely contaminate estimates of the parameters in the factor and sales models (Piatek and Papaspiliopoulos 2018). This paper thus adopts a more flexible approach to factor analysis that relaxes the Gaussian
assumption on the latent factors. That is, it assumes latent factors are non-parametric, emerging from the Dirichlet Process (DP) (West et al. 1994; Escobar and West 1995; Bruce 2019). The DP is a distribution over distribution, which is generated based on the properties of the common Dirichlet distribution (See Bruce 2019 for details and intuition). It provides a flexible, non-parametric approach that will adapt factors to the non-Gaussian shape of the cues measures.

Technically, we model the latent factors $f_{ij}$ to come from an unknown distribution $f_{ij} \sim F$, and that distribution in turn emerges from a Dirichlet Process, $F \sim DP(\hat{\alpha}F_0)$, with precision parameter $\hat{\alpha} > 0$ and prior expectation $F_0 = N(0, I_k)$. Now, suppose we drop the $(i,j)$ subscript to simplify the exposition and define the full set of factors for (say) $p$ observed cues $f = \{f_1, \ldots, f_p\}$; and for any $n=1,\ldots,p$, let $f^{(n)}$ be the set of factors without the factor $f_n$, $f^{(n)} = \{f_1, \ldots, f_{n-1}, f_{n+1}, \ldots, f_p\}$. Then, from Blackwell and Macqueen (1973), the predictive distribution of $f_n$ obtained after integrating over $F$ is the following conditional (discrete) mixture prior:

$$f_n | f^{(n)} = \hat{\alpha}_{n-1} N(f_n | 0, I_k) + (1 - \hat{\alpha}_{n-1}) \sum_{f \neq f_n} \delta_{f_n}(f_n)$$

where $\hat{\alpha}_{n-1} = \hat{\alpha} / (\hat{\alpha} + n - 1)$ and $\delta_{f_n}(f_n)$ is a Dirac delta function. Equation (7) means that conditional on $f^{(n)}$, we sample $f_n$ from a prior normal with probability $\hat{\alpha}_{n-1}$; otherwise, we sample it from one of the previous draws with equal probability (See Bruce 2019 for details).

We describe posterior computations given this mixture prior (Equation 7) and the factor model (Equation 4) in Web Appendix W1; further details and intuition for DPM are available in the above-cited sources. It is essential, however, to appreciate intuition: if we use strict assumptions (e.g., Gaussian densities) when data densities are unknown or unclear, we are likely
to leave some variation in the data undiscovered (e.g., Hjort et al. 2010) and as a result bias our inferences (e.g., loadings parameters). Obviously, the DPM introduces complexity and computational costs to factor analysis, but we gain the flexibility to respond to observed non-normal cues (Carvalho et al. 2008). We next consider the potential endogeneity issues before giving an overview of the estimation procedure and its advantages.

*Controlling for Endogeneity: Advertising and Price*

Advertising and price are potential sources of endogeneity, even though the case for ad spending or content endogeneity in our sample is not as compelling. If managers allocate advertising (spending or content) strategically (e.g., based on sales), advertising might be endogenous; but our estimation relies on weekly data, so endogeneity may not be a major concern (Sethuraman, Tellis, and Briesch 2011). Firms usually determine media schedules for brands in yearly meetings (Leeflang et al. 2000), and based on the performance of individual brands, some minor changes might occur during the year; but media budgets (or campaign schedules) cannot vary within a week. To confirm this, we held interviews with two global media/brand managers, both working for major FMCG firms that own several brands in our dataset, and with a manager at a larger media-planning agency. Results suggest it would be nearly impossible to adjust media budgets on a weekly basis and that, for a number of reasons, the highest frequency with which companies could adjust their budget is monthly. First, TV networks plan and cut commercial breaks several days in advance, so they are unable to accept short-term changes to schedules. Most ad slots (especially for popular shows) sell well in advance (Belch and Belch 2009), such that it is almost impossible for brand managers to purchase a suitable slot at short notice. Cancellation periods usually end six weeks before an ad airs; after this period, companies may increase, but not decrease, their spending levels. Second, the many parties involved (e.g., media-
planning agency, network, advertising company) make it difficult to coordinate changes at short notice. Finally, marketing research companies often supply observed sales metrics one week after their ad spending is determined.\(^8\) Thus, for our weekly dataset, endogeneity of ad spending or content is unlikely to be a concern.

In contrast to advertising, retailers can readily adjust the prices of FMCGs on a daily or weekly basis. A careful empirical analysis would thus account for price endogeneity, which may arise because of unobserved demand shocks (e.g., Besanko, Gupta, and Jain 1998). To control for endogeneity, we adopt an instrumental variable (IV) approach, even though finding such variables (i.e., both uncorrelated with the error term of the sales equation and sufficiently correlated with price) can be difficult. Thus, for IVs, we use the prices of different product categories (e.g., for a yogurt brand, we use the average prices of chocolate bars, shampoos, shower gels, household detergents, and razors as instruments; see “Data” section). The rationale (e.g., Hausman 1997) is that while demand shocks may be uncorrelated between markets, price changes may be correlated across some categories because they share common cost (e.g., labor, ingredients, transportation) structures (e.g., see discussions in Sotgiu and Gielens 2015; Rossi 2014).\(^9\) We test for the strength and the validity of these IVs using the multivariate F-statistic (Angrist and Pischke 2009) and the Sargan test. Results confirm that our IVs are correlated with price \((F\text{-test} >10; \ p\text{-value}=.00)\) but uncorrelated with the error term (Sargan test; \(p\text{-value}= .42\)).

However, a concern for any IV model for weekly price is that firms are unlikely to generate new prices every week; instead, price levels may emerge from changes from a previous time period. Therefore, we also conducted a Durbin-Watson test of the residuals obtained from a regression of price against the IVs; the results show that we cannot reject the autocorrelation of these

---

\(^8\) Promotion schedules are generally even less flexible than media schedules.

\(^9\) We thank the AE for this helpful comment.
residuals for 94% of the brands in our sample \((p\text{-value} < .05)\). Accordingly, we specify the
following model to account for endogeneity:

\[
p_{it} = \theta_{it} + P_{it}^{IV} \hat{\eta} + \nu_{it}^p, \quad \text{and} \\
\theta_{it} = \rho \theta_{it-1} + \nu_{it}^\theta,
\]

where \(\nu_{it}^p \sim N(0, \kappa_i^2), [\nu_{it}^p, w_{0it}] \sim N(0, H_i)\), and \(H = \begin{bmatrix} \kappa_i^2 & \omega_i \\ \omega_i & \sigma_{0i}^2 \end{bmatrix}\).

Thus, Equations 8 and 9 model price across brands as functions of (i) IV covariates \(P_{it}^{IV}\)
or the average weekly prices of other product categories; (ii) normal measurement and system
noises, \(\nu_{it}^p\) and \(\nu_{it}^\theta\), respectively; the former potentially correlated with \(w_{0it}\), the random noise
from the sales (price) equation (1); and (iii) a latent, time-varying component \(\theta_{it}\) governed by an
AR(1) process (e.g., See Bruce, Murthi, and Rao 2017). The latter reflects the persistence of
weekly prices. To control for potential endogeneity (which is relevant when \(\text{Cov}(\nu_{it}^p, w_{0it}) \neq 0\)),
we condition the analysis of Equations 1 and 2 on \(\nu_{it}^p\) (See e.g., Rossi, Allenby, and McCullough
2005).

**MODEL ESTIMATION**

Bayesian inference via MCMC is a natural choice for this study, given the cue data issues
outlined earlier and the need for a unified procedure that can jointly estimate parameters of both
the ad-sales and factor models. Traditional approaches to estimation would likely involve a two-pass
procedure: in the first pass, estimate the loadings and factors; in the second, estimate the ad-
sales model using estimates of the factors as variables. This procedure however suffers an errors-
in-variables problem, because estimates in the second pass ignore factor uncertainty (Geweke
and Zhou 1996).

Fortunately, our estimation accounts for this uncertainty; this involves the iterative
sampling of the posteriors of (i) the time-varying parameters \(p(G_i, \theta_\tau | R_\tau, p_\tau, \Psi_i)\) of the ad-sales
(Equations 1–3) and measurement models (Equations 8–9); and (ii) the parameters
$p(f_{ij}, A | y_{ij}, \beta_{ij})$ of the factor model (Equation 4-7), where $\Psi_i$ is a collection of all fixed brand and campaign parameters, which includes ad effectiveness $\beta_{ij}$. We recover $p(G_u, \theta_u | R_u, p_u, \Psi_i)$, by sampling the conditionals, $p(\theta_u | G_u, \ldots)$ and $p(G_u | \theta_u, \ldots)$ (See Bruce, Murthi and Rao 2017 for algorithm details). For example, conditional on $G_u$ (equation 1), brand sales $R_u$ provide no more information for estimating $\theta_u$; we therefore recover $\theta_u$ from the linear state space model defined by Equations 2, 8, and 9. As a result, we apply the basic Kalman Filter/Smoother Algorithm to estimate $p(\theta_u | G_u, R_u, p_u, \Psi_i)$ and MCMC to its related fixed parameters in $\Psi_i$ (Carter and Kohn 1994). The other conditional distribution $p(G_u | R_u, p_u, \Psi_i)$ is also linear in the goodwill parameter $G_u$, so we can again apply the Kalman Filter algorithm, but with the conditional variance $w_{0ui} | \psi_{ui}$ to control for price endogeneity (see, e.g., Rossi et al. 2005; Bruce, Murthi and Rao 2017). Now, conditional on goodwill $G_u$, we can use basic MCMC ideas to recover the fixed parameters, including brand campaign effectiveness $\beta_{ij}$. Lastly, we recover the joint distribution $p(f_{ij}, A | y_{ij}, \beta_{ij})$ of the main parameter (factors and factor loading matrix), given the sparsity and DP priors in equations 5-7 (See Posterior Computations and Simulations Experiments in Web Appendix W1).

DATA AND IDENTIFICATION

Advertising Model Data

From Nielsen, we obtained weekly retail, scanner panel, and media data for 62 brands and 177 campaigns across six FMCG categories (chocolate bars, yogurt, razors, shampoo,
shower gel, and household detergents\textsuperscript{10}) in the German market; the sample covers an average market share of 63\% per category. It contains, for each brand, weekly sales ($R$) and TV ad spending ($a$); information on several control variables ($z$), such as price and in-store promotions; and ad spending on Internet, billboard, and print, all for 200 weeks from March 2010 to December 2013. Model identification thus draws upon the week-to-week variations in brand sales, ad spending, and controls within this balanced panel (62 brands, 200 weeks). Tables 1-2 provide descriptive statistics.

\textbf{Ad Campaign Data}

The 62 brands in our sample spent 1.6 billion euros across 177 ad campaigns and 325 executions (See Table 3 for a campaign summary). Here, an execution refers to a single creative (TV spot); and a campaign includes all such creatives with a common theme. Several executions belonging to the same ad campaign could differ in minor ways, such as spot length. Furthermore, executions are often aired for very short periods, typically for a few weeks; and because of this data sparsity, we adopt ad campaigns, which last an average duration of 22 weeks, as our level of analysis. Notably, the thematic variation across campaigns helps link executions to campaigns. Moreover, the correlation (co-movement) among the campaign ad spending of the brands in our sample is low (median = 0.061; mean = 0.051); this facilitates the identification of the separate effect of each campaign on brand sales. We now describe branding cues and later, the procedure employed to obtain them for TV ads.

\textbf{---Insert Tables 1–2 here ---}

\textbf{Ad Campaign Data}

The 62 brands in our sample spent 1.6 billion euros across 177 ad campaigns and 325 executions (See Table 3 for a campaign summary). Here, an execution refers to a single creative (TV spot); and a campaign includes all such creatives with a common theme. Several executions belonging to the same ad campaign could differ in minor ways, such as spot length. Furthermore, executions are often aired for very short periods, typically for a few weeks; and because of this data sparsity, we adopt ad campaigns, which last an average duration of 22 weeks, as our level of analysis. Notably, the thematic variation across campaigns helps link executions to campaigns. Moreover, the correlation (co-movement) among the campaign ad spending of the brands in our sample is low (median = 0.061; mean = 0.051); this facilitates the identification of the separate effect of each campaign on brand sales. We now describe branding cues and later, the procedure employed to obtain them for TV ads.

\textbf{---Insert Tables 3 here ---}

\textsuperscript{10} Household detergents include, for example, multi-purpose, glass, or anti-limescale cleaners.
Branding cues. Our most novel data are the salience and association (attributes and benefits) cues extracted from the actual TV ads that correspond to the above spending. From literature related to building brand awareness (Baker, Honea, and Russell 2004; Romaniuk 2009; Stewart and Furse 1986), we adopt the following salience cues: the number of times ads display/mention a brand name, logo, and product; and the duration, in seconds, for which an ad presents a logo and product. Earlier, we outlined several types of attributes and benefits, or association cues (i.e., product-related, non-product-related (price and packaging), functional, experiential, and symbolic) (see, e.g., Keller (1993). For our analysis, we take the number of times ads explicitly or implicitly present these cues. A description of the cues is in Table 4; coding instructions are in Web Appendix W2A. Finally, we code the content of each execution separately; aggregate the results to the campaign level by weighing each execution according to its spending level; and log transform the positive cue values (Cameron and Trivedi 2005, pp. 531) for estimation.

---- Insert Table 4 here ----

Control variables. We also measure two control variables to indicate whether an ad contained an emotional appeal and whether the advertised product is a line extension; both features could influence ad effectiveness (Tellis 2004). Emotional appeal is the maximum value of five commonly used emotions (humor, eroticism, romance, warmth, and nostalgia); we measure it using established, multi-item, seven-point scales (Chattopadhyay and Basu 1990; Edell and Burke 1987). For the line extension variable, we use a dummy variable (1 = line extension; 0 = no line extension). Please see Web Appendix W2B for more information on these control variables.

Coding procedure. Consistent with MacInnis, Rao, and Weiss 2002, we employed independent experts trained to code the content of each ad in terms of the branding cues and two
control variables. The experts, graduate students of a large German university, were regular users of the advertised categories. Groups of two to seven experts coded each variable/cue, depending on the task (e.g., two experts coded whether the product was a line extension, but seven experts coded the emotional appeal). All underwent a two-day training session in which we discussed each variable and clarified any wording problems (see Web Appendix W3). After the training, we provided each expert with a USB stick containing all ads and the coding instructions, so they could code the ads at their own pace, at home. We advised them to code no more than five ads per day and to take a break after watching two ads in a row. The experts generally needed between 25 minutes and two hours to code an ad; coding efficiency improved with experience. The entire coding took four months. The sequence of ads differed for each expert, to avoid order biases. We assessed intercoder reliability using Krippendorff’s (1980) alpha, to ensure the quality of the measurements. All the constructs exceeded the critical value of .67 (see Web Appendix W3).

Table 5 lists the mean, standard deviation, minimum, and maximum for each cue ($y_1$-$y_{17}$) across all 177 campaigns. Note that the minimums are zero for all benefits and attributes cues, and for the brand name awareness cue. That is, even though we (and TV viewers) always observe some ad features (e.g., its appeal), brands generally emphasize different cues across campaigns, depending on (say) campaign objectives: e.g., awareness or image association, or attempts to persuade or inform (e.g., Rossiter and Percy 1998). Thus, the histograms of the cues in Figure 2 show significant clustering at zero (i.e., the data are a discrete-continuous mixture). Recall we model this as a left-censoring problem (using a Type I Tobit) since not accounting for clustering at zero will bias our estimates. Kamakura and Wedel (2001) offered a Tobit factor model for similarly censored marketing data: a sample with information on purchase volumes
across multiple categories. Note too that (in Figure 2) the empirical distributions of cues (even after excluding the discrete zeros) appear to be non-normal. As a result, we find that Jarque-Bera tests for observations (Jarque and Bera 1987) reject hypotheses that branding cues arise from normal distributions.

---- Insert figure 2 and table 5 here ----

RESULTS

Tables 6-10 and Figures 3-8 summarize the main estimates of the proposed models, which include estimates from the ad response model (equations 1-3) and the sparse, NP Tobit factor model (equations 4-7). Again, we take a basic approach to the NP modeling of the cue factors, assuming they emerge from DPMs (e.g., Carvalho et al. 2008; Bruce 2019). Sparsity, on the other hand, allows loadings to be zero (e.g., West 2003; Lucas et al. 2006), so cues (factors) become irrelevant if their loadings in the related rows (columns) in the $\Lambda$ matrix are all zeros. As outlined earlier, we exploit this flexibility to identify the factors while estimating the loadings matrix. That is, we estimate the model (Equations 1-9) separately with 3, 4, and 5 factors, and retain a model with three factors ($k=3$) (See details below). In this case, the matrix $\Lambda$ is sufficiently sparse to obtain unique factor identification without additional (“hard”) restrictions (see, e.g., Kaufmann Schumacher 2019, pp. 117). Results below uncover the relative effects of different branding cues on ad effectiveness (Table 7) from a set of significant loadings $\lambda_{a}$; those with large posterior probabilities of being non-zero, $\pi_{a} = \Pr(\lambda_{a} \neq 0 | y) > .99$ (see Carvalho et al. 2008). The posterior distributions of the factor $\rho_{i}$ and cue specific probabilities $\pi_{a}$ then summarize evidence for sparsity. Significant estimates (highlighted in bold) from the ad response and campaign effectiveness models remain those whose 95% highest probability density interval (HPDI) excludes zeros. Overall, results show that salience and association cues
can positively moderate the effect of TV ads on sales. Brand communication was however most effective when managers combined salience (in particular logo) cues with explicit benefit and attribute cues and applied explicit rather than implicit cues. Finally, we implemented the model with 25000 iterations, accepting every 5th and discarding the first 5000 burn-in iterations; Web Appendix W4 reports preliminary evidence for some key results, and Web Appendix W5 appraises the performance of our factor methods.

We next review results from ad response and factor models, consider the implications of the latter results on cue strategies of select brands, and assess the effects of the observed cues on ad effectiveness.

Advertising Response Model

Table 6 shows the means, standard deviations, and 95% HPDIs for population estimates from the ad response model11. In Table 6, consider the population parameters ($\eta$) for the price, promotion, and communication variables ($z$) in the ad response model. As expected, price has a significant, negative effect on sales (-.98) while both in-store promotions (.28) and competitor price (.11) have significant, positive effects. The effects of other advertising activities (e.g., print, online, and billboard advertising) and competitor advertising are however insignificant. The table also reports that while the effects of the brand intercept ($\alpha_0$) and line extensions ($\mu_1$) are insignificant, the effects of emotional appeal ($\mu_2$) on ad effectiveness are positive and significant at the 90% HPDI level. This is in line with previous studies (MacInnis, Rao, and Weiss 2002) and suggests that, in our setting, ad content has a greater influence on ad effectiveness than (say) the intrinsic differences among brands. Finally, the goodwill intercept

12 See Web Appendix W5: Figure F5 shows the proposed ad response model fits the data quite well. Table T7 reports brand level estimates; Figure F6 displays posterior population densities.
$G_i$, the population mean of initial brand goodwill in the absence of advertising (Naik and Raman 2003), is also significant and positive.

We also report estimates for ad effectiveness and endogeneity. Recall, our analysis controls for the potential endogeneity of price, which in this case means controlling for the co-movement of the measurement $v_i'$ and the observation equation noises $w_{itv}$ for each brand (in Equations 1 and 8, respectively). We find co-movement, $\text{Corr}(w_{itv}, v_i')$ to be significant for 54 of the 62 brands in the sample (See Figure F7, Web Appendix W5); this implies that controlling for price endogeneity was necessary (Bruce et al. 2017). Figure 3 graphs the distribution of mean ad effectiveness ($\beta$) across the 177 campaigns, with most estimates between 0 and .02. We also report estimates of short- and long-term ad elasticities (Figure F8, Web Appendix W5). Overall, these results are in line with estimates from previous studies of FMCG brands (e.g., Srinivasan, Vanhuele, and Pauwels 2010; van Heerde et al. 2013). Later, we consider how specific branding cues moderate long-term ad elasticity.

---- Insert Table 6 and figures 3 ----

**Branding Cue Loadings and Sparsity**

Table 7 and Figures 4-5 summarize the main posterior estimates of the proposed sparse three-factor model. Table 7 reports significant loadings $\lambda_{ch}$ ($\pi_{ch} > .99$) for each cue-factor association, and the effect $\lambda_{0h}$ of each factor $(h)$ on ad effectiveness $\beta_s$, along with the mean factor loading probability $\rho_h$. Notably, merely 2 of the 17 branding cues cross-load: Frequency, Product and Duration Product; thus, we observe sparsity in panel (a) of Figure 4, which shows

\begin{equation}
\frac{\partial G}{\partial a} \frac{a}{G} = \frac{a \beta g'(a)}{(1-\delta)G}
\end{equation}

---

\[^{12}\text{We calculate advertising elasticities as follows: } AD (Spend) \text{ Elasticity, } E_a = \frac{\partial G}{\partial a} \frac{a}{G} = \frac{a \beta g'(a)}{(1-\delta)G} \]
that the distribution of the mean probability that loadings take non-zero values \( \pi_{ch} = \Pr(\lambda_{ch} \neq 0 | y) \) is bimodal, with mostly very small or large values (West 2003; Lucas et al. 2006) on the unit interval. With minimal cross-loadings, large loadings \( \lambda_{ch} \) facilitate the naming/interpretation of the latent factors, the primary objective of sparsity. Similarly, from a visual inspection of the mean loading probabilities \( \rho_h \), (i.e., .31, .20 and .12 for Factors 1, 2, and 3, respectively, in Table 7) and their distributions in panels (b) to (d) (Figure 4), it is clear Factor 1 explains most of the variation among branding cues.

--- Insert table 7 and figure 4 here ---

Moreover, one can interpret factor 1 as the propensity to use salience (S) cues, in particular showing the logo and product in ads. On this dimension (factor 1), salience cues have higher loadings than non-salience cues, more (5) salience cues load on factor 1, and cues Duration Product and Duration Logo have the highest loadings (1.99 and 1.94, respectively). Nonetheless, brand advertisers often incorporate association (attribute-A and Benefits-B) cues such as ingredients or taste, as well as salience cues (e.g., brand name, logo). Thus, as expected, loadings for association cues Explicit Product-Related, Explicit Functional, and Explicit Experiential (0.63, 0.61, and .33, respectively) are also significant. Factor 1 hence captures covariance among different cues, which reflects the cue strategies of our brands. The above loadings suggest brands may explicitly communicate product-related (attributes) with experiential (benefit) cues, as frames 1-3 of Figure 5 indicate. Here, they combine a product cue (e.g., “It contains Ayurveda oil,” frame 1) with those describing the sensory experience of product consumption (“…and thus induces an exotic scent and a nice skin feeling,” frames 2 and 3). Yet what is more notable about this cue strategy is the preference for incorporating cues explicitly rather than implicitly; we return to this point below.
While factor 1 captures most of the co-variation in our cue data (i.e., $\rho_1 = .31$, .20 and .12 for factors 1, 2 and 3, respectively in Table 7), factors 2 and 3 record other brand message strategies in the sample. For example, factor 2 reflects the propensity of some brands to use *Implicit Functional* and *Implicit Symbolic* benefit cues, which have the highest loadings (1.00 and 0.9, respectively). Even so, *Implicit Non-Product-Related: Packaging* (0.64), *Explicit Symbolic* (0.60) and *Explicit Non-Product-Related: Price* (0.35) all load significantly. Thus, managers may at times combine an extrinsic benefit (Explicit symbolic) of a product (“Where does confidence begin?”, frames 4, Figure 5) with an implicit functional one (e.g., “Directly with yourself and the best shave from Gillette,” frame 5) and with Explicit Non-Product-Related cues (“For less than 1 Euro per week,” frame 6). Conversely, factor 3 is determined by *Frequency Product* (0.51), although *Duration Product* (0.27) also loads significantly. These loadings suggest that some brands may focus mostly on salience or awareness cues.

Finally, sparsity not only supports factor interpretation but also helps exclude irrelevant cues by shrinking their loadings to zero, which yields rows of zeros in the factor loadings matrix (e.g., Kaufmann and Schumacher 2019). In our case, four cues did not load significantly onto any factor in our proposed sparse model (i.e., $\pi_{ij} < .95$): *Explicit Non-Product-Related: Packaging, Implicit Non-Product-Related: Price, Implicit Product-Related,* and *Implicit Experiential*. Sparsity excluded the first two, explicit packaging and implicit price messages, because they were rarely used in brand communications (See Figure 2). The latter two, *Implicit Product-Related* and *Implicit Experiential*, are constructs weakly related to all other variables.

--- Insert figure 5 here ---

*Moderating Effect of Branding Cue on Advertising Effectiveness*
So far, we have described factor-cue associations, in particular how cues define the three retained factors and the tendency of brands to combine different cues in their advertising. Yet which factors, and in turn cues, drive the effectiveness of brand advertising? First, Table 7 and Figure 6 (the 95% HPDI plots) show that factor 1 significantly and positively (λ<sub>01</sub> = 0.011) moderates the effect (β) of ad spending (a) on sales (R); but the similar effects of Factors 2 and 3 (λ<sub>02</sub>, λ<sub>03</sub>) are insignificant. Factor 1 reflects brands’ tendencies to use salience cues and to combine these with explicit Product-Related attributes and explicit Functional and Experiential benefit cues. Thus, salience and some association cues have positive effects on TV ads, but brand communication is most effective when managers combine salience cues with explicit benefit and attribute cues.

There is evidence that making salience cues prominent can positively influence ad effectiveness in terms of recall and persuasion (e.g., Stewart and Furse 1986; Book and Schick 1997), which could ultimately affect sales. Furthermore, because implicit cues load on Factor 2, which had no effect on advertising (Table 7), explicit cues are more effective than implicit ones; this, as noted earlier, seems more consistent with our context (TV ads for low-involvement FMCG).

Nonetheless, brand managers and creative talent within ad agencies may wish to know which cues have the greatest impact on the ad-sales relationship. We can obtain some insight by computing the effect of branding cues on ad elasticity (E). We do this by taking the derivative of elasticity with respect to the content loaded on factor 1, to obtain the following expression:

\[
\frac{\partial E_a}{\partial y_c} = \frac{ag'(a)\lambda_{01}\hat{\lambda}_{(c,1)}}{(1-\delta)Gy_c}
\]

where \(g'(a) = 1/(1+a)\), \(\delta\) is the carryover rate; \(\lambda_{01}\) is the effect of factor 1 on campaign effectiveness; \(\hat{\lambda}_{(c,1)}\) is the \(c\)th element from the first row of the matrix \((A^'A)^{-1}A'\); and \(y_c\) is the

---

13 We thank the A/E for this suggestion.
observable branding cue (recall we estimated the model using $ln(y)$, the log of the positive cue values; Cameron and Trivedi 2005, pp. 531). As a result, equation (10) captures a positive but diminishing effect of cues on advertising (e.g., Teixeira, Wedel, and Pieters 2010). Table 8 shows that duration logo has the largest effect on long-term ad elasticity (0.0196, $p$-value < 0.001). In other words, a 1 second increase on average has a 1.9% incremental effect on long-term elasticity. As expected, the second highest effect then comes from frequency logo, which has a .5% effect on elasticity; finally, the third highest is from product duration (.32%). The other effects (0.08% to .14%) are considerably smaller.

--- Insert Figure 6 and Table 8 here ---

Analyzing the Brand Communication Strategies of Different Brands

We innovate in this paper to introduce Bayesian methods for a non-Gaussian, sparse Tobit factor model. Our approach is a “hybrid” that combines the parsimony of confirmatory factor models while retaining some of the flexibility of exploratory factor models (Lu, Chow, and Loken 2016). Thus, the method provided interpretable (“cleaner”) loadings, which earlier helped rank the effectiveness of different branding cues. The method, by zeroing a subset of these loadings, can also uniquely identify three factors from non-normal branding cues. We now use these three factors to analyze and compare the brand communication strategies of different brands.

Specifically, we plot the first and second factors, which cover most of the seventeen branding cues, for each campaign in two dimensions. Note that we rescaled the factors to a 10-point scale for visual clarity. The resulting plots help reveal potential brand communication issues and support comparisons of brand communications across competitors. For illustration, we consider two plots. In the first, we plot campaigns for the yogurt category; that with the largest
total sales volume (Figure 7). Brand names are anonymized into numerical form. Each circle in Figure 7 represents a different campaign, and the circle size reflects the brand’s average sales level. The second plot (Figure 8) includes a sample of some of the largest and smallest brands in the data (based on a median split of the average sales value per category).

--- Insert figures 7 and 8 here ---

From the diagnostic plots, we see that brand communication strategies vary significantly across campaigns and within brands. Because factor 1 drives advertising effectiveness, marketers should strive to be in the upper quadrants; brands located in the lower quadrants (e.g., brand 5 and 2 in Figure 7) might thus want to revisit their brand communication strategies. For example, brand 5 displays the logo and product in its advertising campaigns for 2 to 5 seconds, which is well below the category average of 8 seconds and explicitly promotes only one product attribute. Our results suggest that brand 5 can enhance its advertising effectiveness by a) making their brand and product more salient and b) having a stronger focus on explicitly stated product attributes in combination with experiential or functional benefits. Other brands, mainly large ones, air multiple campaigns to focus on different aspects (e.g., brand 1). Brand 2, however, appears to employ a very similar branding strategy across its multiple campaigns. It might enhance its advertising effectiveness by focusing on the branding cues for the first factor (e.g., display the logo and product for a longer time or integrate explicitly mentioned product attributes) in at least some of these campaigns. Note that no yogurt brand strongly concentrates on branding cues that identify the second factor (e.g., non-product-related attributes or implicit benefits), maybe because managers know that these branding cues are less effective.

In Figure 8, we find that most small brands focus less on factor 1 (e.g., brands 7, 8, 10, 11). Further examining the specific cues reveals that they concentrate particularly little on branding
cues pertaining to salience, despite the fact that increasing awareness is critical for small, lesser-known brands (Elliott and Percy 2007). Brands 7 and 8, for instance, mention their brand name just once and display their logo for only 4 and 5 seconds, respectively. Furthermore, brands 10 and 11 strongly focus on the second factor by including symbolic benefits and packaging cues in their advertising, while our results suggest that they should instead focus on factor 1 and explicitly mention product attributes, functional benefits, or experiential benefits. This result could reflect the fact that most of the small brands in our sample are niche brands, which may try to differentiate their advertising message by focusing on less commonly employed brand associations, such as symbolic benefits or non-product-related attributes.

**Factor Selection and Non-Gaussian Factors**

As a final step, we assess the appropriateness of the proposed three-factor model and the impact of the non-parametric (DP) assumption. First, to gauge the number of factors, we estimated an exploratory PCA (principal component analysis) independently of the ad-sales model. The procedure recovered three ($k=3$) latent components from the 17 branding cues, where factors 1, 2 and 3 explained 58%, 26% and 14% (respectively) of the explained variance. Here, however, we jointly estimate the ad-sales and the NP factor model for $k=2$ up to 5 factors to assess the appropriate number, then re-estimate the best-fitting model, assuming factors were Gaussian (See previous estimation algorithm). Table 9 compares the alternative models based on the deviance information criterion (DIC). The DIC penalizes gains in fit that come solely from model complexity in Bayesian hierarchical models, for which the number of parameters is unclear (Spiegelhalter et al. 2002). Since our primary uncertainty is the factor model, the DIC here is based on the conditional likelihood $l(y|\Lambda, \beta)$. The proposed non-parametric, sparse model (NP, $k=3$) outperforms all alternative models as indicated by its DIC values. Further,
specifications with $k=4$ and 5 yield 1 and 2 columns, respectively, in the factor matrix $A$ with zero loadings (See Figure F9, Web Appendix W5). Notably, the proposed outperformed the Gaussian factor model (model 5). Lastly, we quantified the negative consequences of the Gaussian assumption on the loadings by comparing the significant samples $\hat{\lambda}_{ch}(\hat{\tau}_{ch} > .99)$ for loadings obtained from the NP ($k=3$) and Gaussian ($k=3$) models; 52.94% of the mean loadings were different at the 95% level (see Simulations in Web Appendix W1 for more details).

---- Insert table 9 here ----

MANAGERIAL IMPLICATIONS

Simulation

We summarize the implications of our results with a simulation. Specifically, we solve a problem that reallocates branding cues to improve each brand’s expected sales. Afterwards, we compare plots of the current and improved ad content strategies, based on the model (P1).

Formally, we solve P1:

$$\max_{y_{11}, y_{12}, \ldots, y_{ij}} \sum_{t=1}^{T} E(R_{it} | G_{i-1})$$

$$st, 0 \leq y_{ij} \leq y_{ij}^m, c = 1,2, \ldots, 5; \hat{y} = \sum_{c=1}^{C} y_{sc}$$

where $\hat{y} = \sum_{c=1}^{C} y_{sc}$ is the actual number of brand association cues in the ad. Thus, we allow for a reallocation across brand association cues while keeping the number of brand associations constant. The measure $y_{ij}^m$ represents the maximum for the salience cues. To determine the maximum for the duration cues (duration logo and duration product), we assume that the period for which an ad can display the logo or product is bounded by spot length. We also restrict the frequency variables (i.e., frequency of brand name, logo, and product) to the sample maximum for all campaigns (observations) that used a similar spot length (e.g., upper limit for a ten second
spot is the maximum value for all ten second spots in the sample). Lastly, \( E(R_{it} \mid G_{it-1}) \) is the one-step-ahead sales \( (R_{it}) \) forecast, and \( y_{oi} \) is the branding cue \( c \) of brand \( i \) and campaign \( j \).

The model-based solution for PI provides new allocations for the branding cues across the respective campaigns; these generated an average 2.3% increase in sales, with increases ranging from 1046€ to over a million euros. Because factor 2 did not moderate ad effectiveness (Table 7), the reallocation proposes a shift from \( i \) symbolic benefits, implicitly stated functional benefits, and non-product-related attributes to \( ii \) explicitly stated functional or experiential benefits and product attributes, as well as a stronger emphasis on salience (e.g., displaying the logo for a longer duration). From the model-based re-allocations, we can also derive new factor scores and graphically compare them to the original scores implied by the cue data. For illustration purposes, we use the new scores from two campaigns for different brands: one that performed poorly on factor 1 and one that performed well. As expected, the model-based factor scores of all campaigns move to the upper left quadrant of the plot (Figure 9). While brand communication changes only marginally for a campaign that already performed well (brand 6), it changes substantially for a poorly performing campaign (brand 10).

---- Insert figure 9 here ----

Finally, the cues for brand 10 reveal that its campaign mentioned the brand name merely once and displayed the product and the logo for just a second. It is unlikely that consumers could have noticed the brand from such a fleeting ad exposure. Notice that the cues for brand 10 were mainly symbolic benefits and non-product-related attributes; our above model-based results, however, suggest that brand 10 should shift to explicit product-related attributes, functional cues, or experiential cues. Naturally, such analyses can help brand and ad agency managers track ad performance to uncover ways to jointly improve ad creative and sales.
CONCLUSIONS AND LIMITATIONS

This paper explored how brand messages or cues embedded in TV advertising influence the effects of ad spending on brand sales. To accomplish this, we first build a dynamic ad response model that quantifies the effects of TV ads on FMCG brand sales, then innovate to build a non-parametric (robust), sparse factor model to extract a set of common factors and interpretable loadings from multiple (non-Gaussian, left-censored) branding cues. The latent factors are assumed to moderate the effects of TV ads on sales, and the loading parameters measure the relative importance of each branding cue to the interpretation of each factor, and in turn to the success of advertising. To obtain interpretable loadings, we adopt Bayesian sparsity for the estimation. The approach is a “hybrid” that combines the parsimony of confirmatory factor models while retaining some of the flexibility of exploratory ones. Again, we take a basic approach to the non-parametric modeling of cue factors, assuming they emerge from DPMs. We calibrate the ad response model with panel and media data from the Nielsen Company for 62 brands and 177 campaigns across six product categories sold in the German market; for the factor model, we obtained branding cues from a set of trained experts who coded all 177 advertising campaigns in terms of 17 cues.

We found a number of substantive results. First, salience and some association cues have positive effects on TV ads, though brand communication is most effective when managers combine salience cues with explicit benefit and attribute cues. Second, because implicit cues load on factor 2, which had no effect on advertising, explicit cues are more effective than implicit ones in our study. Third, we find that duration logo had the largest effect on long-term ad elasticity (0.0196, p-value < 0.001). In other words, a 1 second increase on average has a 1.9% incremental effect on long-term elasticity. As expected, the second highest effect comes from
frequency logo, which had a .5% effect on elasticity, with product duration having the third-highest effect. Fourth, brand advertisers often combine explicit product attributes with explicit experiential benefits, which can then positively affect sales. This strategy (e.g., combining product-related cues with the sensory benefit cues) may be effective because it gives consumers “reasons” to accept benefit claims. Our simulation then suggests that improving brand communication within ads can yield on average 2.3% increase in sales. Finally, brand and ad agency managers can use these methods and findings to identify strategies for improving creative ad content.

The study has several limitations that future work could address. First, some of its findings may not generalize to other contexts. In particular, its findings arise from established FMCG brands, and the relevant (say) benefit cues might differ across product categories. As an example, symbolic cues, which correlate with prestige, are likely more important for luxury items. Second, the study focuses on 17 branding cues, but there might be other, equally important cues for brands (e.g., slogan, jingles, brand character). Still, we excluded similar alternatives from our data because they had extremely low variance (i.e., almost all ads use slogans, and virtually none of them employ brand characters).

REFERENCES


### TABLE 1: OPERATIONALIZATION OF TIME-SERIES DATA

<table>
<thead>
<tr>
<th>Variable</th>
<th>Operationalization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volume sales ($R$)</td>
<td>Sales in kg&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Advertising spending ($a$)</td>
<td>Gross TV advertising spending (in €)</td>
</tr>
<tr>
<td>Price ($z_1$)</td>
<td>Price per kg (in €)</td>
</tr>
<tr>
<td>In-store promotions ($z_2$)</td>
<td>% of stores having an in-store promotion&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Competitor price ($z_3$)</td>
<td>Market-share weighted competitor price per kg&lt;sup&gt;c&lt;/sup&gt; (in €)</td>
</tr>
<tr>
<td>Other advertising activities ($z_4$)</td>
<td>Gross spending on other advertising activities (in €) (billboard, internet, and print)</td>
</tr>
<tr>
<td>Competitor advertising ($z_5$)</td>
<td>Total competitor gross TV advertising spending (in €)</td>
</tr>
</tbody>
</table>

<sup>a</sup> Volume sales indicates sales in kg for the yogurt, chocolate bars, shampoo, shower gel, razors, and household detergents category. For the razor category, it represents the number of razors per pack.<sup>b</sup> In-store promotions indicates the percentage of stores that feature brand i. Note that because not all stores have the same consumer traffic, the percentage is weighted by store size.<sup>c</sup> Market share calculations include in-store brands.

### TABLE 2: DATA SUMMARY, TIME-SERIES DATA

<table>
<thead>
<tr>
<th>Category</th>
<th>Number of brands in dataset</th>
<th>Combined value market share</th>
<th>Volume sales in kg&lt;sup&gt;e&lt;/sup&gt;</th>
<th>Price per kg (€)</th>
<th>% of stores having an in-store promotion</th>
<th>Other advertising activities&lt;sup&gt;e&lt;/sup&gt; (€)</th>
<th>Competitor price per kg (€)</th>
<th>Competitor advertising (€)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>62</td>
<td>63%</td>
<td>157,914</td>
<td>10.56</td>
<td>5%</td>
<td>9,429</td>
<td>9.96</td>
<td>952,892</td>
</tr>
<tr>
<td>Yogurt</td>
<td>15</td>
<td>57%</td>
<td>365,779</td>
<td>3.08</td>
<td>6%</td>
<td>3,627</td>
<td>2.86</td>
<td>660,491</td>
</tr>
<tr>
<td>Chocolate bars</td>
<td>14</td>
<td>66%</td>
<td>125,853</td>
<td>9.32</td>
<td>5%</td>
<td>10,948</td>
<td>8.79</td>
<td>1,915,673</td>
</tr>
<tr>
<td>Shampoo</td>
<td>10</td>
<td>66%</td>
<td>105,805</td>
<td>10.42</td>
<td>8%</td>
<td>16,753</td>
<td>9.09</td>
<td>1,079,530</td>
</tr>
<tr>
<td>Shower Gel</td>
<td>8</td>
<td>64%</td>
<td>158,145</td>
<td>6.55</td>
<td>8%</td>
<td>6,622</td>
<td>5.75</td>
<td>254,709</td>
</tr>
<tr>
<td>Razors</td>
<td>6</td>
<td>61%</td>
<td>19,193</td>
<td>7.10</td>
<td>2%</td>
<td>18,853</td>
<td>7.42</td>
<td>314,218</td>
</tr>
<tr>
<td>Household detergents</td>
<td>9</td>
<td>61%</td>
<td>122,502</td>
<td>3.20</td>
<td>4%</td>
<td>4,808</td>
<td>3.01</td>
<td>848,082</td>
</tr>
</tbody>
</table>

<sup>e</sup> Volume sales indicates sales in kg for the yogurt, chocolate bars, shampoo, shower gel, razors, and household detergents category. For the razor category, it represents the number of razors per pack.<sup>e</sup> Other advertising activities includes gross spending for internet, billboard, and print advertising.

### TABLE 3: DATA SUMMARY, ADVERTISEMENT CAMPAIGN DATA

<table>
<thead>
<tr>
<th>Category</th>
<th>Advertising spending per week</th>
<th># of advertisement campaigns per brand</th>
<th>Total # of campaigns</th>
<th># of executions per brand</th>
<th>Total # of executions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>125,165</td>
<td>0</td>
<td>2,073,480</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Yogurt</td>
<td>125,113</td>
<td>0</td>
<td>2,073,480</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Chocolate Bars</td>
<td>178,115</td>
<td>0</td>
<td>1,504,102</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Shampoo</td>
<td>138,616</td>
<td>0</td>
<td>1,893,780</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Shower Gel</td>
<td>43,763</td>
<td>0</td>
<td>1,019,420</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Razors</td>
<td>123,313</td>
<td>0</td>
<td>1,328,400</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Household detergents</td>
<td>101,528</td>
<td>0</td>
<td>1,106,625</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

<sup>e</sup> Advertisement campaigns aired for an average of 22 weeks.
<table>
<thead>
<tr>
<th>Observable Branding Cues</th>
<th>Variable Type</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>Count</td>
<td>Weighted average(^a) number of times the brand name was mentioned</td>
</tr>
<tr>
<td>Frequency</td>
<td>Count</td>
<td>Weighted average(^a) number of times the logo was displayed</td>
</tr>
<tr>
<td>Frequency</td>
<td>Count</td>
<td>Weighted average(^a) number of times the product was shown</td>
</tr>
<tr>
<td>Duration</td>
<td>Count</td>
<td>Weighted average(^a) length of time the logo was displayed (in seconds)</td>
</tr>
<tr>
<td>Duration</td>
<td>Count</td>
<td>Weighted average(^a) length of time the product was shown (in seconds)</td>
</tr>
<tr>
<td>Explicit</td>
<td>Count</td>
<td>Weighted average(^a) number of explicitly mentioned, product-related attribute cues (e.g., ingredients)</td>
</tr>
<tr>
<td>Explicit</td>
<td>Count</td>
<td>Weighted average(^a) number of explicitly mentioned, non-product-related attribute cues, related to price</td>
</tr>
<tr>
<td>Explicit</td>
<td>Count</td>
<td>Weighted average(^a) number of explicitly mentioned, non-product-related attribute cues, related to packaging</td>
</tr>
<tr>
<td>Implicit</td>
<td>Count</td>
<td>Weighted average(^a) number of implicitly mentioned, product-related attribute cues (e.g., ingredients)</td>
</tr>
<tr>
<td>Implicit</td>
<td>Count</td>
<td>Weighted average(^a) number of implicitly mentioned, non-product-related attribute cues, related to price</td>
</tr>
<tr>
<td>Implicit</td>
<td>Count</td>
<td>Weighted average(^a) number of implicitly mentioned, non-product-related attribute cues, related to packaging</td>
</tr>
<tr>
<td>Explicit</td>
<td>Count</td>
<td>Weighted average(^a) number of explicitly mentioned, functional benefit cues (e.g., cleans, removes dandruff, stills hunger)</td>
</tr>
<tr>
<td>Implicit</td>
<td>Count</td>
<td>Weighted average(^a) number of implicitly mentioned, functional benefit cues (e.g., cleans, removes dandruff, stills hunger)</td>
</tr>
<tr>
<td>Explicit</td>
<td>Count</td>
<td>Weighted average(^a) number of explicitly mentioned, experiential benefit cues (e.g., odor, taste, haptics)</td>
</tr>
<tr>
<td>Explicit</td>
<td>Count</td>
<td>Weighted average(^a) number of explicitly mentioned, symbolic benefit cues (e.g., prestige, makes one feel accepted, increases social approval)</td>
</tr>
<tr>
<td>Implicit</td>
<td>Count</td>
<td>Weighted average(^a) number of implicitly mentioned, functional benefit cues (e.g., cleans, removes dandruff, stills hunger)</td>
</tr>
<tr>
<td>Implicit</td>
<td>Count</td>
<td>Weighted average(^a) number of implicitly mentioned, experiential benefit cues (e.g., odor, taste, haptics)</td>
</tr>
<tr>
<td>Implicit</td>
<td>Count</td>
<td>Weighted average(^a) number of implicitly mentioned, symbolic benefit cues (e.g., prestige, makes one feel accepted, increases social approval)</td>
</tr>
<tr>
<td>Control</td>
<td>Percentage</td>
<td>Indicates for each campaign the percentage of executions that promote a new line extension.</td>
</tr>
<tr>
<td>Control</td>
<td>Interval (1–7 Likert scale)</td>
<td>Formative construct based on several multi-item scales (entertainment, humor, erotic, surprise, warmth, nostalgia, romance) indicating how emotional (high arousal) the ad execution is. We take the weighted average(^e) of all executions belonging to the same campaign.</td>
</tr>
</tbody>
</table>

\(^a\) Notes: We take the averages of all executions belonging to the same ad campaign and weight each execution according to its spending level.
### TABLE 5: DESCRIPTIVE STATISTICS FOR BRANDING CUES

<table>
<thead>
<tr>
<th>Observable branding cue</th>
<th>Mean</th>
<th>SD</th>
<th>Max.</th>
<th>Min.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brand Name (y₁)</td>
<td>2.38</td>
<td>1.15</td>
<td>6.92</td>
<td>.00</td>
</tr>
<tr>
<td>Logo (y₂)</td>
<td>3.33</td>
<td>1.45</td>
<td>9.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Product (y₃)</td>
<td>3.17</td>
<td>1.44</td>
<td>9.11</td>
<td>1.00</td>
</tr>
<tr>
<td>Duration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logo (y₄)</td>
<td>7.41</td>
<td>4.27</td>
<td>23.64</td>
<td>2.00</td>
</tr>
<tr>
<td>Product (y₅)</td>
<td>7.95</td>
<td>3.80</td>
<td>21.28</td>
<td>1.00</td>
</tr>
<tr>
<td>Explicit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Product-Related (y₆)</td>
<td>1.26</td>
<td>.95</td>
<td>4.00</td>
<td>.00</td>
</tr>
<tr>
<td>Non-Product-Related: Price (y₇)</td>
<td>.13</td>
<td>.28</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Non-Product-Related: Packaging (y₈)</td>
<td>.04</td>
<td>.16</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Implicit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Product-Related: Price (y₉)</td>
<td>.37</td>
<td>.63</td>
<td>4.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Non-Product-Related: Packaging (y₁₀)</td>
<td>.01</td>
<td>.11</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Explicit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Functional (y₁₁)</td>
<td>.95</td>
<td>1.20</td>
<td>6.00</td>
<td>.00</td>
</tr>
<tr>
<td>Experiential (y₁₂)</td>
<td>.78</td>
<td>.79</td>
<td>4.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Symbolic</td>
<td>.43</td>
<td>.58</td>
<td>2.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Implicit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Functional (y₁₃)</td>
<td>.21</td>
<td>.61</td>
<td>6.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Experiential (y₁₄)</td>
<td>.28</td>
<td>.55</td>
<td>3.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Symbolic</td>
<td>.14</td>
<td>.34</td>
<td>1.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Notes: Descriptive statistics are displayed at campaign level. Campaign data are based on the weighted average of the executional data.

### TABLE 6: CONTROL VARIABLES AND PARAMETERS

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>STD</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ad Sales Model</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Price (z₁)</td>
<td>-.98</td>
<td>.26</td>
<td>-1.43</td>
<td>-.57</td>
</tr>
<tr>
<td>In-store promotions (z₂)</td>
<td>.28</td>
<td>.05</td>
<td>.20</td>
<td>.37</td>
</tr>
<tr>
<td>Competitor price (z₃)</td>
<td>.11</td>
<td>.02</td>
<td>.07</td>
<td>.24</td>
</tr>
<tr>
<td>Other advertising activities (z₄)</td>
<td>.21e-2</td>
<td>.23e-2</td>
<td>-.18e-2</td>
<td>.59e-2</td>
</tr>
<tr>
<td>Competitor advertising (z₅)</td>
<td>.59e-2</td>
<td>.34e-2</td>
<td>-.19e-2</td>
<td>.14e-1</td>
</tr>
<tr>
<td>Campaign Effectiveness Model</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brand Intercept, α₀i</td>
<td>.04e-2</td>
<td>.05e-2</td>
<td>-.74e-2</td>
<td>.85e-2</td>
</tr>
<tr>
<td>Line Extension, X₁</td>
<td>.06e-2</td>
<td>.16e-2</td>
<td>-.26e-2</td>
<td>14e-2</td>
</tr>
<tr>
<td>Emotional Appeal, X₂**</td>
<td>.20e-2</td>
<td>.15e-2</td>
<td>.01e-2</td>
<td>.40e-2</td>
</tr>
<tr>
<td>Brand Goodwill Model</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goodwill Intercept, Gᵢ</td>
<td>8.17</td>
<td>1.93</td>
<td>5.02</td>
<td>11.31</td>
</tr>
</tbody>
</table>

Notes: Bolded cells indicate 95% HPDI. **90% HPDI, 2.5% and 97.5% Percentiles
### TABLE 7: Brand Cues Loaded on Factors 1, 2, and 3 (Non-Gaussian Factors)

<table>
<thead>
<tr>
<th>Observable Branding Cues</th>
<th>Type</th>
<th>$\hat{\lambda}_{ch}$</th>
<th>$\rho$</th>
<th>$\hat{\lambda}_{ch}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Factor 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration Product ($y_2$)</td>
<td>S</td>
<td>1.99</td>
<td>.31</td>
<td>.011</td>
</tr>
<tr>
<td>Duration Logo ($y_4$)</td>
<td>S</td>
<td>1.94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency Logo ($y_2$)</td>
<td>S</td>
<td>1.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency Product ($y_3$)</td>
<td>S</td>
<td>1.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency Brand Name ($y_1$)</td>
<td>S</td>
<td>1.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Explicit Product-Related ($y_6$)</td>
<td>A</td>
<td>0.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Explicit Functional ($y_{12}$)</td>
<td>B</td>
<td>0.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Explicit Experiential ($y_{13}$)</td>
<td>B</td>
<td>0.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Factor 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Implicit Functional ($y_{15}$)</td>
<td>B</td>
<td>1.00</td>
<td>.20</td>
<td>ns</td>
</tr>
<tr>
<td>Implicit Symbolic ($y_{17}$)</td>
<td>B</td>
<td>0.90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Implicit Non-Product-Related: Packaging ($y_{11}$)</td>
<td>A</td>
<td>0.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Explicit Symbolic ($y_{14}$)</td>
<td>B</td>
<td>0.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Explicit Non-Product-Related: Price ($y_7$)</td>
<td>A</td>
<td>0.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Factor 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency Product ($y_3$)</td>
<td>S</td>
<td>0.51</td>
<td>.12</td>
<td>ns</td>
</tr>
<tr>
<td>Duration Product ($y_5$)</td>
<td>S</td>
<td>0.27</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** Bolded cells indicate loadings, $\hat{\lambda}_{0h}$ and $\hat{\lambda}_{ch}$ for which cues specific probability $\pi_{ch} > .99$; ns - not significant.


### TABLE 8: CHANGE IN LONG TERM AD ELASTICITY PER UNIT CHANGE IN AD CUES

<table>
<thead>
<tr>
<th>Observable Brand Cues</th>
<th>Mean</th>
<th>STD</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency Brand Name ($y_1$)</td>
<td>0.0014</td>
<td>0.0321</td>
<td>0.0008</td>
<td>0.0115</td>
</tr>
<tr>
<td>Frequency Logo ($y_2$)</td>
<td>0.0053</td>
<td>0.0322</td>
<td>0.0009</td>
<td>0.0120</td>
</tr>
<tr>
<td>Frequency Product ($y_3$)</td>
<td>0.0014</td>
<td>0.0169</td>
<td>0.0002</td>
<td>0.0027</td>
</tr>
<tr>
<td>Duration Logo ($y_4$)</td>
<td>0.0196</td>
<td>0.0646</td>
<td>0.0036</td>
<td>0.0531</td>
</tr>
<tr>
<td>Duration Product ($y_5$)</td>
<td>0.0032</td>
<td>0.0257</td>
<td>0.0006</td>
<td>0.0068</td>
</tr>
<tr>
<td>Explicit Product-Related ($y_6$)</td>
<td>0.0013</td>
<td>0.0165</td>
<td>0.0002</td>
<td>0.0024</td>
</tr>
<tr>
<td>Explicit Functional ($y_{12}$)</td>
<td>0.0014</td>
<td>0.0182</td>
<td>0.0002</td>
<td>0.0026</td>
</tr>
<tr>
<td>Explicit Experiential ($y_{13}$)</td>
<td>0.0008</td>
<td>0.0113</td>
<td>0.0001</td>
<td>0.0014</td>
</tr>
</tbody>
</table>

**Notes:** Bolded cells indicate 95% HPDI. 2.5% and 97.5% Percentiles.

### TABLE 9: PERFORMANCE OF ALTERNATIVE MODELS

<table>
<thead>
<tr>
<th>Models</th>
<th>LL</th>
<th>DIC</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. NP Factors ($k=2$)</td>
<td>-1308.1</td>
<td>4913.6</td>
<td>2</td>
</tr>
<tr>
<td>2. NP Factors ($k=3$)</td>
<td>-1303.4</td>
<td>4844.8</td>
<td>1</td>
</tr>
<tr>
<td>3. NP Factors ($k=4$)</td>
<td>-1292.4</td>
<td>4993.1</td>
<td>3</td>
</tr>
<tr>
<td>4. NP Factors ($k=5$)</td>
<td>-1193.4</td>
<td>5298.2</td>
<td>5</td>
</tr>
<tr>
<td>5. Gaussian Factors ($k=3$)</td>
<td>-1819.3</td>
<td>5001.3</td>
<td>4</td>
</tr>
</tbody>
</table>

**Notes:** NP – Non-parametric; DIC = deviance information criterion. LL = log likelihood.
FIGURE 1: CONCEPTUAL MODEL

FIGURE 2: HISTOGRAM OF BRANDING CUES
FIGURE 3: HISTOGRAM - MEAN AD CAMPAIGN EFFECTIVENESS, $\beta$

FIGURE 4: HISTOGRAM OF FACTOR LOADING PROBABILITIES

(a) $\pi_{\theta_1}$

(b) $\pi_{\theta_2}$

(c) $\pi_{\theta_3}$

(d) $\pi_{\theta_4}$
FIGURE 5: EXAMPLES OF COMBINATIONS OF BRAND ASSOCIATION CUES IN ADVERTISEMENTS

1. Palmolive Ayuritvel shower gel
2. With Indian Ayurveda oil
3. Improves your well-being and revives your body and mind

4. Where does self-confidence begin?
5. Directly with yourself and the best shave from Gillette
6. For less than 1 Euro per week

FIGURE 6: PLOT OF FACTOR 1 BY AD CAMPAIGN (Factor 1 Mean and 95% HPDI)
FIGURE 7: TWO-DIMENSIONAL PLOT FOR YOGURT CATEGORY

Notes: The numbers denote the different brands; each circle represents one campaign. The circle size represents the brand’s average weekly sales value.

FIGURE 8: TWO-DIMENSIONAL PLOT FOR SMALL VS. LARGE BRANDS

Notes: The numbers denote the different brands; each circle represents one campaign. The circle size represents the brand size (small or large, based on a median split of the average weekly sales value).

FIGURE 9: ILLUSTRATION OF MODEL-BASED RESULTS

Notes: The numbers denote the different brands; each circle represents one campaign. The striped circles (“mb”) indicate the model-based positioning of the campaigns.