Abstract

We propose a theoretical model of advertising content, based on an information-persuasion trade-off, where the “persuasive” content of an advertisement is defined as the content that is not objective information. Brands must allocate available time in an advertisement between imparting objective information about the product’s attributes and other persuasive methods. Our theoretical model gives rise to a structural ordered probit model, which we estimate with data from the OTC Analgesics Industry. We test four predictions of the model.

We find that stronger vertical differentiation is positively associated with the delivery of more product information in a brand’s advertisements: Brands with higher levels of quality (on each of the quality dimensions for which we have data) will include more information cues in their advertisements. We also find that comparative advertisements contain significantly more product information relative to noncomparative advertisements. Brands with higher market shares and brands competing against generic substitutes with higher market shares have less information content in their advertising. The proposed method of measuring and analyzing information content of advertising extends and improves upon existing techniques for measuring advertising information.

Keywords: information content, advertising, information-persuasion tradeoff, content analysis
Information Content of Advertising: Theory and Empirical Evidence

How much information brands choose to disclose in advertisements is a question of considerable theoretical and empirical debate. Recent research in marketing and economics provides some theoretical predictions on the relationship between market structure or firm size and the amount of information transmitted (Anderson and Renault 2009; Guo and Zhao 2009; Sun 2010). The empirical work on advertising content is split into two camps with essentially no overlap. The first camp uses Resnik and Stern’s (1977; henceforth RS) methodology to analyze content (for a summary, see Abernethy and Franke 1996). The second camp treats advertising content as a choice variable (Anderson et al. 2010; Bertrand et al. 2010; Liaukonyte 2010). This paper develops an information-persuasion trade-off theory model, where the “persuasive” content of an ad is defined as the content that is not objective information, and enables us to bridge the gap between the two camps by applying the theory model to the novel data.

The key premise of this paper is that there exists an optimal amount of objective information to be included in an advertisement. Advertisements that provide too little objective information about the brand arguably waste the opportunity to sufficiently convince prospective consumers to buy it (Jacoby 1977). Conversely, those that provide too much information may crowd the ad message and lead to information overload for the consumer (Chervany and Dickson 1974; Ansari and Mela 2003). Factors such as motivation and the ability to process information (Cacioppo and Petty 1985; MacInnis and Jaworski 1989) mediate individual responses to advertising. Therefore, the complexity of advertising, including too many information cues, can create attention wear-out (Pieters, Rosbergen, and Wedel 1999), which suggests that an optimal amount of information content exists. The optimal amount of information content may vary systematically across brands within an industry and may be partially explained by observable
factors such as brand type, brand size, suitability of various combinations of information, and recent news about the product.

We formulate a theoretical model in which firms decide how much objective information to include in an advertisement. The model allows for a trade-off between objective information and other persuasion, subject to random factors intrinsic to specific advertisements. The amount of time spent on information—as approximated by the number of information cues—has systematic and random components. This approach allows us to provide a theoretical underpinning for an ordered probit model of the number of information cues in advertisements. Furthermore, the theoretical model also yields predictions on the relationship between information content and various observable performance measures (such as market share).

The empirical component of this paper examines the relationship between market variables and the information content of advertising. First, we classify and fully measure different types of advertising content within an entire industry. Second, we relate the extent of information disclosure to the market share of a brand, its core vertical characteristics, and the market share of the generic substitute. To accomplish this, we formulate four hypotheses about the influence of the type of advertising, brand vertical characteristics, market share, and generic substitute market size on the information content of advertising.

The empirical testing of the formulated theory focuses on the OTC analgesics industry, for several reasons. First, television advertising constitutes a large fraction of sales, implying that it is the most important marketing strategy used by the industry to communicate to its consumer. Second, the products sold by firms differ significantly and discretely in their characteristics, so that there is a range of meaningful information to potentially communicate (Lancaster 1971, Christou and Vettas 2008). Third, the relevant product information conveyed through advertisements is concentrated in experience- and credence-based characteristics, rather than
search characteristics that consumers can learn in the store before purchase (Erdem, Keane and Sun 2008). Fourth, the industry has an oligopolistic market structure, which is most likely to deliver the strongest results, as opposed to a monopoly (where comparisons of behavior across firms with different attributes would not be possible) or competitive environments (where advertising levels tend to be relatively low) (Sutton 2007). Fifth, product differentiation, both real and spurious, is important because we emphasize the trade-off between persuasion and information, with the optimal mix depending on a product’s characteristics. For all of these reasons, the US OTC analgesics industry provides an ideal test-bed for the theory.

In line with our theoretical predictions stemming from a trade-off between informative and persuasive components of advertising content, we find empirical support for the hypothesis that stronger vertical differentiation is positively associated with greater information content in advertisements: Brands with higher levels of quality (on each of the quality dimensions for which we have data) will include more information cues in their advertisements. We also find that comparative advertisements contain significantly more information than noncomparative advertisements. Furthermore, we find that larger brands and a higher market share of the generic version of a brand are both associated with less informative ads. We show that the information content of advertising is a function of a brand’s characteristics and that the decision to include a comparative statement is endogenous. Significant estimation bias results from not controlling for the endogeneity of the decision to use comparative advertising and from the endogeneity of market share. Finally, we show that the results are largely robust to using the RS information content definition; however, the RS methodology does not fully account for all of the aspects of information that are important to consumers.
**Background and Definitions**

In general, information in advertisements can be (1) about the brand, brand attributes, benefits, users, or usage situation; (2) cognitive, emotional or subconscious; and (3) context information, including consumers’ past experience (MacInnis, Moorman, and Jaworski 2001; Vakratsas and Ambler 1999). In our theoretical model and then later in the empirical application, we focus on the first aspect of brand information (observable, objective information cues) and treat cognitive, emotional, and contextual information as the unobservable, subjective element of an ad, which we define as the *persuasive component*.\(^1\) Figure 1 summarizes our approach. The shaded area represents the observable informative component of an ad, whereas the non-shaded area represents unobservable (to us) persuasive component.

Contrary to the classic content analysis (see meta-analysis of Abernethy and Franke 1996), which summarizes information cues within advertisements and uses univariate analysis to compare scenarios, we use an ordered probit model to study the *determinants* of the distribution of cues. The ordered probit is directly derived from a theoretical model that describes the equilibrium choice of information content as a function of the exogenous variables. Our model is based on an intuitive relationship: a greater amount of information content is associated with a larger number of information cues.

It is difficult to derive clear-cut testable hypotheses based on existing theory. First, most economic and marketing theories of advertising do not address the content of advertising (cf. Anderson and Renault 2006, 2009) and most models are monopoly models: even oligopoly models usually assume that firms are symmetric. Second, standard advertising models typically

---

\(^1\) The persuasive component can also be traceable in well-defined studies (e.g., Maheswaran and Joan Meyers-Levy 1990), but it is outside of the scope of this paper.
address only one type of advertising (persuasive or informative), while any given advertisement likely incorporates several components simultaneously.

Figure 1. Determinants of Information Disclosure

![Diagram of Information Cues in Advertising and Determinants of Information Disclosure]

**The Model**

Our theoretical model distinguishes between two types of advertisement content: information and persuasion. The firm faces a trade-off: given the limited amount of time available in an ad, the firm must decide how much of that time to use to provide information and how much of it to use to persuade consumers to buy the brand through channels other than objective information. We assume that the persuasive power of an advertisement depends on the number of seconds devoted to the persuasive component, and a random term which reflects idiosyncratic features of a particular advertisement. Let the persuasive power of the ad be

\[ P(s, \varepsilon) = (\bar{p} - \varepsilon)s \]
where $s$ is the number of seconds given to persuasion, $\bar{p}$ is a constant and $\varepsilon$ is a random term.\textsuperscript{2} Panel A of Figure 2 illustrates the function $P(s, \varepsilon)$. The x-axis indicates the number of seconds used for persuasion. As more time is used for persuasion, the function $P(s, \varepsilon)$ increases from the right to left. The y-axis shows the total benefit from persuasion. The more time is used for persuasion, the larger the total benefit.

The marginal persuasion for an advertisement with $s$ seconds of persuasion and a given draw of the random term $\varepsilon$ is therefore $\bar{p} - \varepsilon$. The persuasion function is linear and therefore the marginal persuasion does not change with the share of the advertisement that is used for persuasion. Panel B of Figure 2 illustrates the function $\bar{p} - \varepsilon$. The x-axis again indicates the number of seconds used for persuasion. Because $\bar{p} - \varepsilon$ does not change with $s$, the function $\bar{p} - \varepsilon$ is parallel to the x-axis. The y-axis depicts the marginal benefit of persuasion. In the econometric model presented later, $\bar{p}$ varies according to observable features of the advertised brand (e.g., market share, observable quality, generic competition, etc.) and of the advertisement itself, such as whether it is a comparative or noncomparative advertisement.

Let each information cue take $\bar{s}$ seconds to convey, so that if there are $S$ seconds in the ad (i.e. $n$ information cues are conveyed), there are $s=S-n\bar{s}$ seconds of persuasion. Let $I_i$ be the benefit of the $i^{th}$ information cue $i$, with $i=1,...,n$. We rank the cues from the highest to lowest information benefit for each given advertisement (and the ranking may differ across advertisements according to the particular theme of the ad). The brand will choose to include the cues delivering the highest information benefit, i.e., those cues for which the values of $I_i$ are the highest.

Since there is a given amount of time, $S$, available in one advertisement, the brand must decide how much of that time to use to provide information and how much to use for persuasion.

\textsuperscript{2} $P(s, \varepsilon)$ does not have to be linear in $s$. It should be an increasing and concave function of $s$. The linear specification simplifies the exposition significantly.
Such trade-off faced by the brand and the total benefit of information is depicted in Panel A of Figure 2. Panel B of Figure 2, on the other hand, shows the marginal benefit of information, which is decreasing in the amount of information already provided. The marginal benefit of information is also a step function. The firm chooses \( s \) (or, alternatively, \( n \)) to maximize the sum of the total benefit of persuasion and the total benefit of information. Formally, the firm solves:

\[
\max_s D(\sum_{i=1}^n I_i + P(s, \varepsilon)) \\
\text{s.t. } n = \frac{S - s}{\bar{s}}.
\]

Here \( D(\cdot) \) is an increasing function representing the firm’s demand as a function of information and persuasion content of an advertisement. The solution to this optimization problem can be described by comparing the incremental benefit from adding a cue to the advertisement to the opportunity cost of reducing the time spent on persuasion. If the advertisement contains \( n-1 \) cues, then the extra benefit from an \( n \)th cue is \( I_n \). We can see this graphically in Panel A of Figure 2. There we see that \( I_4 \), the marginal benefit of the fourth information cue is larger than \( I_5 \), the marginal benefit of the fifth information cue, while the slope of the persuasion function, \( \bar{\rho} - \varepsilon \), is such that:

\[
I_5 < \bar{s}(\bar{\rho} - \varepsilon) < I_4.
\]

This implies that the optimal number of cues depicted in Panel A of Figure 2 is four.

Panel B of Figure 2 shows this solution in a way that allows us to introduce our statistical model in a straightforward way. Define for all \( I_i \) the value \( \varepsilon_i \) such that

\[
I_i = (\bar{\rho} - \varepsilon_i) \bar{s},
\]

so that \( \varepsilon_i \) is the threshold value of the random error such that the firm chooses to include at least \( i \) information cues in the advertisement if \( \varepsilon > \varepsilon_i \). In Figure 2, Panel A example we have depicted \( \varepsilon_4 < \varepsilon < \varepsilon_5 \); hence, in this particular illustration, the brand chooses to include four information cues in the advertisement.
The formulation above underpins our statistical model. Accounting for the fact that there can be no fewer than 0 cues or no more than \( C = S / \bar{S} \) we have the following mapping from the value of \( \epsilon \) to the number of cues, here denoted by \( y \):

\[
\begin{align*}
y &= 0, \text{ for } \epsilon \leq \epsilon_1 \\
... \\
y &= i, \text{ for } \epsilon_i < \epsilon \leq \epsilon_{i+1} \\
... \\
y &= n, \text{ for } \epsilon > \epsilon_n.
\end{align*}
\]

The basic intuition of this statistical model is the following: When the (negative) random shock is very small (\( \epsilon < \epsilon_1 \)), which implies that an advertisement has ability to have very strong persuasive power (relative to the benefit of information in that particular advertisement), then the firm has no incentive to include information cues. As the persuasive power of an advertisement gets smaller, the firm chooses to include more information cues.

From the specification above we can construct a probability distribution of observing the corresponding number of cues, where \( F(\cdot) \) denotes the cumulative distribution of \( \epsilon \):

\[
\begin{align*}
y &= 0, \text{ with probability } F(\epsilon_1) \\
... \\
y &= i, \text{ with probability } F(\epsilon_{i+1}) - F(\epsilon_i) \\
... \\
y &= n, \text{ with probability } 1 - F(\epsilon_n).
\end{align*}
\]

If \( F(.) \) is normally distributed, and corresponds to an ordered probit model.\(^3\) In particular, we show how the probability function in the model above can be written as the textbook version of the ordered probit, where \( \epsilon_i = \alpha_i - X\beta \) and therefore \( F(\epsilon_i) = \Phi(\alpha_i - X\beta) \). In this ordered probit model specification the unobserved components, i.e. \( \epsilon \), are drawn from a normal distribution, and the cutoff values \( (\alpha_i) \) are such that the realization of a latent variable (explained

\(^3\) For a comprehensive review of ordered models, see Greene 1997 and Woolridge 2001.
component plus noise) lies within a range that corresponds to each specific number of cues. $\beta$ is a $K \times 1$ vector of parameters, and $X$ is a $K \times 1$ vector of observable features of the brand, which does not include a constant. For a positive $\beta$, an increase in $X$ will lower the threshold $\varepsilon_i$ which will in turn make adding additional information cue more desirable. Therefore, in this model, for a positive $\beta$, larger $Xs$ are associated with more information content in an advertisement.

We can rewrite $F(\varepsilon_i) = \Phi(\alpha_i - X\beta)$. First, recall that $\varepsilon_i = \bar{p} - \frac{I_i}{\bar{s}}$. Here $\bar{p}$ is a variable, the value of which determines the benefit of persuasion content. Hence, we set $\alpha_i - X\beta = \bar{p} - \bar{I} - \frac{I_i}{\bar{s}}$. It is useful to extract a constant from $-\frac{I_i}{\bar{s}}$ and to rewrite the equality as $\alpha_i - X\beta = \bar{p} - \bar{I} - \frac{I_i}{\bar{s}}$. Then $X\beta = \bar{I} - \bar{p}$ so that, consistent with the discussion in the preceding paragraph, the variables $X$ increase the benefit of information, or, equivalently, they decrease the benefit of persuasion.

Second, we can define the cutoff value $\alpha_i = -\frac{I_i}{\bar{s}}$, which determines whether the firm is including $i$ or $i+1$ information cues. The econometric model does not identify the constant term $\bar{I}$ separately from the cutoffs $\alpha_i$, which we need to keep in mind when we interpret the cutoffs.

In our framework, the cutoffs have a clear structural interpretation. In particular:

$$\alpha_{i+1} - \alpha_i = \frac{I_i - I_{i+1}}{\bar{s}}.$$

Thus, conditional on $\bar{s}$, which is unobserved, differences in the cutoffs provide information on differences in the information benefits of an additional cue. This observation enables us to take econometric model results and use them to draw a graph of the estimated cutoffs and compare it directly with theoretical relationship depicted in Figure 2, which we implement in the empirical section below.

---

4 For concreteness, we describe the random term as entering the persuasion power, but it could just as well enter the marginal information benefit described below. Or, indeed, the random term could enter on both sides and the analysis captures the net effect.
Figure 2. Tradeoff Between Informative and Persuasive Content

Panel A

- Information Benefit
- Persuasion Benefit
- Incremental Benefit of 4th Information Cue
- Incremental Benefit of 5th Information Cue
- Opportunity Cost of 5th Information Cue
- $P(s, e) = s(\bar{p} - \varepsilon)$

Panel B

- Incremental Information Benefit
- Incremental Persuasion Benefit
- $\bar{p} - \varepsilon_1$
- $\bar{p} - \varepsilon_2$
- $\bar{p} - \varepsilon_3$
- $\frac{\partial P}{\partial s}(s, e) = \bar{p} - \varepsilon$

Length of an Ad

Length of Time for 1 Cue

Optimal Number of Cues

Time Spent on Informative Cues Included in an Ad

Persuasive Content
We estimate this structural model and use it to examine the relationships between fundamental variables, entering in $X$ (e.g., a market share of a firm), and the number of information cues that a firm includes in an advertisement. Before discussing the relationships of interest we state and prove the main theoretical result, which drives the rest of our analysis and helps us formulate key hypotheses.

**Lemma 1.** An increase in $I_i$ stochastically increases the number of information cues.

**Proof.** From the analysis above, $i$ cues will be advertised if $\epsilon \in (\epsilon_i, \epsilon_{i+1})$, where, as above, $\epsilon_i = \bar{\epsilon} - \frac{l_i}{s}$. The corresponding probability of observing $i$ cues is $P_i = F(\epsilon_{i+1}) - F(\epsilon_i)$, or

$$P_i = F(p - \frac{l_{i+1}}{s}) - F\left(p - \frac{l_i}{s}\right).$$

Suppose now that $I_i$ increases while retaining its position as the $i$th largest information benefit. Then $P_i$ increases at the expense of $P_{i-1}$, while all other probabilities remain unchanged. Hence the number of cues increases stochastically. Now suppose that the increase in $I_i$ raises it to the $j$th highest cue, with $j < i$. Then each intervening cue is promoted so that the probability of observing at least that number of cues rises. The probability of observing $i+1$ cues or more stays the same, as does the probability of observing each number less than $j$. Again, the number of cues stochastically increases. QED

**Vertical Characteristics**

First, we consider intrinsic characteristics of a product. In our empirical application of the theoretical model the examples of intrinsic attributes include but are not limited to: strength of pain relief, relative efficiency, safety, etc. In our analyzed industry, these variables are naturally exogenous to the information decision because they depend on the medical properties of the active ingredients in analgesic pain relievers, which in turn are regulated by the Federal Drug
Administration (FDA). Therefore, the causality direction can be clearly identified, and we can investigate how different locations in the product characteristics space are associated with information disclosure.

The driving idea here is that the information benefit to the brand from communicating a characteristic is larger the stronger the performance of the brand in that characteristic. Communicating a weak characteristic does not give as much incremental benefit as communicating a strong one. Characteristics which are important to consumers are more likely to be communicated, and additional strength in any particular cue will raise its relative information benefit and also make it more likely to be included in the advertisement.

Thus, using the result from Lemma 1 we formulate:

\[ H_1: \text{Brands with higher levels of quality (on each of the quality dimensions for which we have data) will include more information cues in their advertisements.} \]

**Comparative Advertising**

The second relationship that we study is the one between the decision to make a comparative claim and amount of information provided. There is no existing theory that tells us whether comparative advertisements should include more or less information than noncomparative advertisements.\(^1\)

A comparative advertisement has less purely persuasive content than a noncomparative ad because the advertisement needs to mention a competitor, and this dilutes the persuasion. The notion is that whatever information a brand chooses to communicate to its consumer, it will need

---

\(^1\) Each comparative advertisement must include at least one cue, since the comparison is made on at least one characteristic. We compare all noncomparative advertisements to all comparative advertisements, conditional on an advertisement having at least one cue.
to mention the competitor, which highlights and reminds the consumer of the competitor’s existence.

On the other hand, there is a potential for more information in comparing two brands than just purporting one brand. A brand might provide more information by saying A is better than B than by saying A is good. Because when a brand mentions that A is better than B, the information that not only A is good, but also that it is better than one of its competitors is communicated to a consumer. Therefore, for the same amount of advertising seconds, brands will provide more information with comparative advertisements due to higher benefit of information associated with comparative advertisements. In other words, we can imagine comparative advertisements shifting the information function.

We might expect a comparative advertisement to both increase the marginal information and decrease marginal persuasion benefit of an advertisement. Both effects cause the number of cues to rise stochastically (see Figure 2).

First, more information is conveyed by comparing two brands than just promoting a brand. Comparing relative performance gives more precise and concrete reference point. On the other hand, an advertisement with comparative content is likely to have a weaker persuasion effect, even if the amount of time devoted to persuasion is the same. This might happen because mentioning the other brand dilutes the persuasion because it reiterates the existence of the rival brand. Previous research (Chou, Franke, and Wilcox 1987; Harmon, Razzouk, and Stern 1983) has found that comparative advertisements have more information, and we expect to find similar patterns. This leads us to formulate our second hypothesis:

**H2:** Comparative advertisements contain more information than noncomparative advertisements.
Market Share

The larger brands benefit less than smaller brands from providing information, because they are already well known and have higher advertising goodwill and brand equity (Simon 1993, Dekimpe and Hanssens 1995). In other words, the incremental benefit is smaller because consumers are already more aware of features of commonly used products.

Another theoretical underpinning to this hypothesis draws from Anderson and Renault (2009), who model advertising as revealing the direct match between consumers and horizontal product characteristics. The model predicts that smaller brands have more incentives to advertise informatively to differentiate themselves from larger rivals and to carve out a niche market of consumers. Conversely, larger brands might not provide information because doing so informs consumers that the brand may not be the best match. Therefore, larger firms prefer to enhance their brands through noninformative rather than informative advertising. They might benefit from broader advertising themes, while smaller brands tend to target niche consumers with more detailed information about the product (see Iyer, Soberman, and Villas-Boas 2005). We therefore expect (and it this is a direct implication from Lemma 1) that the advertisements for larger brands will have less informative content than those for smaller brands. We formulate our third hypothesis in the following way:

H3: Larger brands (as measured by market share) have less information content in their advertisements than smaller brands.

Generics

An important characteristic of our analyzed market and many other consumer product categories is the presence of generic substitutes, and we extend our theoretical model to allow
their existence. If the market consists of branded and generic products, then each generic product corresponds to a particular brand, i.e. is the closest substitute for that brand. Thus informative advertisements can also increase demand for the generic products as long as consumer is aware that there are a lot of shared attributes between the branded product and its generic counterpart.\(^2\)

Thus, brands that emphasize product quality also provide free advertising for generic products. We argue that the larger the generic counterpart, the less brands want to emphasize shared qualities. For example, consider an advertisement that contains partly persuasive and partly informational content. The persuasive part creates subjective differentiation and improves perceived quality only of the branded (i.e., advertised) product; the informative part increases demand for both the generic substitute and the advertised brand. Because both persuasive and informative themes have diminishing returns (for a review of advertising response functions, see Bagwell 2007), the proportion of informative to persuasive advertising should differ depending on the detrimental effect of information spillover. Thus, the larger the market share of the generic substitute, the less the brand gains from emphasizing its active ingredient–based product through informative advertising. This implies that the marginal incentive for informative content is lower for brands with large market shares of their generic counterparts.

We capture these effects through an extension of the theoretical analysis described above. In the model so far, with no generics, we postulated that the increase in demand for a brand would be \(D(\sum_{i=1}^{n} l_i + P(s, \varepsilon))\). Now, in the presence of a generic, suppose that the demand increase due to advertising for the branded good is subject to some leakage to the generic products. The leakage is assumed to be increasing in the generic market share and in the information content, so that the generic products get more of the customers diverted from branded products, the larger the generic's market share and the larger the information content of

\(^2\) In our analyzed market, OTC analgesics, branded products tend to be equivalent in quality to their generic counterparts, due to regulatory environment that this market is facing.
an advertisement. Furthermore, if the incremental leakage from adding an information cue is increasing in the market share of generic products, then the information content will be lower the higher the share of the generic products is. With these stipulations in mind, we formulate our fourth hypothesis:

**H4:** *Brands with a higher market share of generic substitute will have less informative advertising content than brands with lower generic market share.*

Next, we describe the data and identification strategy that we use in empirical application of the above presented theoretical model.

**Data and Content Analysis**

We use sales and advertising data from the OTC analgesics industry in the United States to evaluate the hypotheses constructed in the theoretical section. The OTC analgesics market covers pain relief medications with four major active chemical ingredients: aspirin, acetaminophen, ibuprofen, and naproxen sodium. The nationally advertised brands include Tylenol (acetaminophen), Advil (ibuprofen), Motrin (ibuprofen), Aleve (naproxen sodium), Bayer (aspirin or combination), and Excedrin (acetaminophen or combination).

---

3 More formally, suppose that the demand increases for the branded good $D\left(\sum_{i=1}^{n} I_i + P(s, \varepsilon)\right) - L\left(s_g, \sum_{i=1}^{n} I_i\right)$ where $D\left(\sum_{i=1}^{n} I_i + P(s, \varepsilon)\right)$ denotes the demand increase to both the branded and the generic counterpart, and $L\left(s_g, \sum_{i=1}^{n} I_i\right)$ denotes the leakage to the generic. The leakage is assumed to be increasing in the generic market share, $s_g$, and in the information content, $\sum_{i=1}^{n} I_i$. Note that $n$ information cues will be preferred to $n+1$ if $D\left(\sum_{i=1}^{n+1} I_i + P(s, \varepsilon)\right) - L\left(s_g, \sum_{i=1}^{n+1} I_i\right) > D\left(\sum_{i=1}^{n} I_i + P(s, \varepsilon)\right) - L\left(s_g, \sum_{i=1}^{n} I_i\right)$ where we have let the argument $n$ denote the use of the first $n$ information cues. Rewriting this condition as $L\left(s_g, \sum_{i=1}^{n+1} I_i\right) - L\left(s_g, \sum_{i=1}^{n} I_i\right) > D\left(\sum_{i=1}^{n+1} I_i + P(s, \varepsilon)\right) - D\left(\sum_{i=1}^{n} I_i + P(s, \varepsilon)\right)$ and noting that the RHS does not depend upon $s_g$, then including the lower number of cues, $n$, is more likely to be preferred the larger is the LHS. That holds if the incremental leakage is increasing in $s_g$. This is a natural condition given that the leakage itself is increasing in $s_g$ (for example, if leakage were proportional to generic size times an increasing and concave function of information cues).
Brands in this category spend larger amounts of money on advertising than in other industries. As reported in Table 1, advertising-to-sales ratios for OTC analgesics typically range from 20% to 30% of sales, making them one of the most heavily promoted manufactured goods. Moreover, brands use both noncomparative and comparative advertising, which allows us to differentiate information content by these two types of advertisements. Finally, the type of cues mentioned (e.g., “strong”) are clearly identifiable, which enables us to avoid making any subjective judgments while coding the information cues.

The advertising data come from TNS Media Intelligence and cover the entire U.S. OTC analgesics product category. The data set contains video files of all advertisements, as well as monthly advertising expenditures, for each product advertised in the OTC analgesics category from 2001 to 2005. The advertising numbers also include expenditures on other media, but almost all the advertising budgets (approximately 90%) were spent on broadcast television advertising, including network and cable television. In our analysis, we examine only the television advertising data.

We watched 4503 individual commercials broadcast during the 2001–2005 period. Each individual advertisement was usually shown multiple times.

The widely used RS method for measuring advertising information categorizes the information provided in advertisements into 14 distinct “information cues,” including price, quality, performance, components, availability, special offers, taste, nutrition, packaging, warranties, safety, independent research, company research, and new ideas. More than 60 studies have applied the RS approach to measure the information content of advertising in different media (Chou, Franke, and Wilcox 1987; Harmon, Razzouk, and Stern 1983; Stern and Resnik 1991), countries (Hong, Muderrisoglu, and Zinkhan 1987; Madden, Caballero, and Matsukubo 1986), and product categories (Stern, Krugman, and Resnik 1981). The results have varied
markedly, even within the same medium, because of the lack of a multivariate statistical analysis, redundant or too broad definitions of information cues, and small sample size (Abernethy and Franke 1996). The main advantage of the RS classification system is the general nature of the information cue categories, which allow for comparison of products from multiple industries. However, this main advantage can also be interpreted as a disadvantage: Categorizing advertising information content into the coarse categories inevitably omits some information consumers might find important. For example, in the OTC analgesics industry, two distinct information cues (e.g., fast and strong) would be coded as one "performance" cue in the RS classification system.

Our attribute coding approach documents every attribute mentioned. For each advertisement, we recorded whether the commercial had any comparative claims and, if so, the specific claim (e.g., faster, stronger). We also noted all information cues mentioned, including the purpose of the drug (e.g., menstrual pain, arthritis, headache), drug efficiency (e.g., strength, speed), safety, and other characteristics. The type of information cues that were mentioned (e.g., “strong”) are clearly identifiable, which enables us to avoid making any subjective judgments while coding the information content. The disadvantage of this approach is that it is industry specific. In discussing our results, we focus on the attribute coding approach, but for robustness, we also present the results using the RS methodology.

Our analysis also incorporates data on strength of pain relief, relative efficiency, and safety for each brand. We collected this information from peer-reviewed medical journals. Clinically, all four main active ingredients have varying degrees of side effects. Since individuals react to each ingredient differently, clinical pain researchers hesitate to assign superiority to any single drug. Each active ingredient has a comparative advantage. Aspirin (brand name: Bayer) is

---

4 As it is standard with the coding of advertising content, we also had independent coders cross-checking our coding.
weak in pain relief but has low, almost nonexistent cardiovascular risk. Naproxen sodium (Aleve) is the most potent drug but is associated with very high gastrointestinal risk. Acetaminophen (Tylenol and Excedrin) has low gastrointestinal risk but is weak in pain relief and has a medium cardiovascular risk. Ibuprofen (Advil and Motrin) and naproxen sodium–based brands (Aleve) have the highest cardiovascular risk but are also the fastest in pain relief.

We quantify or rank all the “true” characteristics that were used in advertising associated with each active ingredient as follows: First, we interpret “fast” as the time taken to achieve perceptible or meaningful pain relief (medical literature terms this “onset to perceptible pain relief”). Second, we interpret claims such as “long lasting” as the duration of meaningful pain relief. Third, we interpret claims about strength (e.g., “strong,” “stronger,” “tougher on pain”) as the maximum level of pain relief achieved; we use the number-needed-to-treat (NNT) measure to approximate analgesic efficiency claims. The NNT is a standard efficiency measure used in pain relief evaluation literature. In Appendix A, we explain how NNT, cardiovascular risk, and gastrointestinal risk are calculated. In addition to using absolute risk and efficacy measures, we supplement the data with relative performance metric for speed of pain relief (see Appendix A). Each active ingredient can be unambiguously ranked by the onset of pain relief using the results in published medical studies.\footnote{We also have information on the relative performance on the duration of pain relief (i.e., "long lasting"; see Appendix A); however, we exclude this metric from further multivariate analysis because it is highly correlated with the NNT measure (correlation coefficient = .94).} The longevity measure is inversely related to the maximum number of regular strength pills allowed within a 24-hour period. Naproxen sodium tops the list in this regard, while acetaminophen has the lowest duration of pain relief. Thus, the information content of advertising is important because it helps consumers match a particular drug to a particular active ingredient.
Table 1 compiles spending and cue information by brand under attribute and RS coding approaches. The first two columns give each brand’s (expenditure weighted) average number of attributes and whether it is comparative. The Aleve and Advil advertisements have slightly more cues than those of Tylenol, Motrin, Excedrin, and Bayer. As Table 1 shows, almost half the advertising expenditures were on comparative advertisements; we pay special attention to the difference in advertising content according to whether an advertisement was comparative or noncomparative. The breakdown is striking across brands: Almost all advertisements for Aleve are comparative, and two-thirds of the Advil advertisements are comparative. For the remaining brands, only about one-third of the advertisements are comparative, except for Excedrin, for which just one-sixth are.

Table 1. Descriptive Statistics of Information Disclosure and Ad Spending

<table>
<thead>
<tr>
<th>Brand</th>
<th>Number of Cues</th>
<th>Number of RS Cues</th>
<th>Comparative?</th>
<th>Avg Monthly Spending per Ad Aired</th>
<th>Average Monthly Sales</th>
<th>Total Ad Spending</th>
<th>Total Sales</th>
<th>Ad to Sales Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advil</td>
<td>3.6 (1.004)</td>
<td>2.59 (0.767)</td>
<td>0.74 (0.441)</td>
<td>$0.14 (0.241)</td>
<td>$23.92 (1.693)</td>
<td>$293.10</td>
<td>$1,374</td>
<td>21.30%</td>
</tr>
<tr>
<td>Aleve</td>
<td>3.77 (1.156)</td>
<td>1.929 (0.548)</td>
<td>0.9 (0.298)</td>
<td>$0.12 (0.293)</td>
<td>$11.41 (1.123)</td>
<td>$174.80</td>
<td>$659</td>
<td>26.50%</td>
</tr>
<tr>
<td>Bayer</td>
<td>3.19 (1.320)</td>
<td>2.229 (0.543)</td>
<td>0.31 (0.461)</td>
<td>$0.10 (0.222)</td>
<td>$7.95 (0.964)</td>
<td>$131.20</td>
<td>$458</td>
<td>28.80%</td>
</tr>
<tr>
<td>Excedrin</td>
<td>2.4 (0.695)</td>
<td>2.139 (0.347)</td>
<td>0.15 (0.359)</td>
<td>$0.26 (0.456)</td>
<td>$12.39 (1.172)</td>
<td>$182.40</td>
<td>$689</td>
<td>26.50%</td>
</tr>
<tr>
<td>Motrin</td>
<td>2.61 (0.937)</td>
<td>1.557 (0.497)</td>
<td>0.37 (0.484)</td>
<td>$0.10 (0.240)</td>
<td>$8.03 (0.762)</td>
<td>$102.00</td>
<td>$466</td>
<td>21.90%</td>
</tr>
<tr>
<td>Tylenol</td>
<td>2.54 (0.957)</td>
<td>2.082 (0.719)</td>
<td>0.28 (0.449)</td>
<td>$0.13 (0.346)</td>
<td>$40.59 (3.195)</td>
<td>$414.90</td>
<td>$2,328</td>
<td>17.80%</td>
</tr>
<tr>
<td>Overall</td>
<td>2.99 (1.143)</td>
<td>2.137 (0.709)</td>
<td>0.46 (0.498)</td>
<td>$0.13 (0.305)</td>
<td>$22.86 (13.677)</td>
<td>$1,299.20</td>
<td>$5,975</td>
<td>21.70%</td>
</tr>
</tbody>
</table>

The fourth column of Table 1 reports the average dollars spent per ad creative, averaged over the advertisements aired in a month. Excedrin spent the most per advertisement per month, with an average of $255,000. Other brands spent between $100,000 and $140,000. These
numbers reflect Excedrin’s reliance on a relatively small number of advertisements in its portfolio at any given time. Although Excedrin spends more per advertisement in any month, it ranks only third in overall advertising spending. During the five-year period analyzed, Tylenol spent the most on advertising ($414 million), constituting 32% of the (dollar weighted) observations. Advil spent around a third less, and the other four brands spent roughly half that amount. The average monthly sales (averaged across brands) are approximately $23 million.

Figure 3. Advertised Attributes and Expenditures

The advertising-to-sales ratios are very high in this industry. As the last column of Table 1 shows, they range from 17.8% for Tylenol to 28.8% for Bayer.

During our analyzed period, 30 different product attributes were mentioned. Figure 3 represents the top 23 attributes and shows the advertising expenditures (in millions of dollars)
spent on advertising those attributes during the sample period.\(^6\) We separate advertising expenditures by the type of advertisement (comparative versus noncomparative). The attributes “fast,” “strong,” “long lasting,” and “trust and/or safety” are among the top five most heavily advertised attributes. These attributes are directly related to the inherent (exogenous) chemical characteristics of each active ingredient in each analyzed brand.

Table 2. Matrix of Frequency of Attributes that are Mentioned Together

<table>
<thead>
<tr>
<th></th>
<th>Fast</th>
<th>Strong</th>
<th>Headache</th>
<th>Long lasting</th>
<th>Safe</th>
<th>Arthritis</th>
<th>Dr. recomm</th>
<th>Liquid gels</th>
<th>Legs/muscle</th>
<th>Gentle on stomach</th>
<th>Back</th>
<th>Fewer pills</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fast</td>
<td>$251.61</td>
<td>$296.52</td>
<td>$70.25</td>
<td>$32.76</td>
<td>$23.80</td>
<td>$16.69</td>
<td>$204.55</td>
<td>$101.42</td>
<td>$66.32</td>
<td>$25.65</td>
<td>0</td>
<td>$613.52</td>
<td></td>
</tr>
<tr>
<td></td>
<td>49.32%</td>
<td>78.63%</td>
<td>23.22%</td>
<td>11.17%</td>
<td>8.77%</td>
<td>6.70%</td>
<td>98.90%</td>
<td>49.52%</td>
<td>56.98%</td>
<td>22.11%</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strong</td>
<td>$251.61</td>
<td>$126.43</td>
<td>$103.37</td>
<td>$93.88</td>
<td>$123.06</td>
<td>$97.80</td>
<td>$140.66</td>
<td>$133.31</td>
<td>$45.46</td>
<td>$51.94</td>
<td>$53.01</td>
<td>$510.20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>41.01%</td>
<td>33.53%</td>
<td>20.26%</td>
<td>12.02%</td>
<td>45.34%</td>
<td>39.28%</td>
<td>68.01%</td>
<td>65.10%</td>
<td>39.05%</td>
<td>44.78%</td>
<td>66.81%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Headache</td>
<td>$296.52</td>
<td>$126.43</td>
<td>$6.22</td>
<td>$28.85</td>
<td>$11.95</td>
<td>$25.42</td>
<td>$84.88</td>
<td>$17.53</td>
<td>$24.40</td>
<td>$15.88</td>
<td>$6.22</td>
<td>$377.09</td>
<td></td>
</tr>
<tr>
<td></td>
<td>48.33%</td>
<td>24.78%</td>
<td>2.06%</td>
<td>9.84%</td>
<td>4.40%</td>
<td>10.21%</td>
<td>41.04%</td>
<td>8.56%</td>
<td>20.96%</td>
<td>13.69%</td>
<td>5.49%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long lasting</td>
<td>$70.25</td>
<td>$103.37</td>
<td>$6.22</td>
<td>$68.50</td>
<td>$153.97</td>
<td>$83.96</td>
<td>$14.43</td>
<td>$64.17</td>
<td>$23.18</td>
<td>$44.99</td>
<td>$100.06</td>
<td>$302.56</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11.45%</td>
<td>20.26%</td>
<td>1.65%</td>
<td>23.36%</td>
<td>56.73%</td>
<td>33.72%</td>
<td>6.98%</td>
<td>31.34%</td>
<td>19.91%</td>
<td>38.79%</td>
<td>88.35%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Safe</td>
<td>$32.76</td>
<td>$93.88</td>
<td>$28.85</td>
<td>$68.50</td>
<td>$115.84</td>
<td>$77.64</td>
<td>$5.51</td>
<td>$29.96</td>
<td>$55.50</td>
<td>$39.29</td>
<td>$30.46</td>
<td>$293.20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.34%</td>
<td>18.40%</td>
<td>7.65%</td>
<td>22.64%</td>
<td>42.68%</td>
<td>31.18%</td>
<td>2.67%</td>
<td>14.63%</td>
<td>47.68%</td>
<td>33.87%</td>
<td>26.90%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arthritis</td>
<td>$23.80</td>
<td>$123.06</td>
<td>$11.95</td>
<td>$153.97</td>
<td>$115.84</td>
<td>$125.59</td>
<td>$21.58</td>
<td>$18.38</td>
<td>$55.21</td>
<td>$19.22</td>
<td>$80.12</td>
<td>$271.42</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.88%</td>
<td>24.12%</td>
<td>3.17%</td>
<td>50.89%</td>
<td>50.43%</td>
<td>50.43%</td>
<td>10.43%</td>
<td>8.97%</td>
<td>47.43%</td>
<td>16.57%</td>
<td>70.75%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dr. recomm.</td>
<td>$16.69</td>
<td>$97.80</td>
<td>$25.42</td>
<td>$83.96</td>
<td>$77.64</td>
<td>$125.59</td>
<td>0</td>
<td>$23.88</td>
<td>$38.35</td>
<td>$4.71</td>
<td>$50.65</td>
<td>$249.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.72%</td>
<td>19.17%</td>
<td>6.74%</td>
<td>27.75%</td>
<td>26.48%</td>
<td>46.27%</td>
<td>0</td>
<td>11.66%</td>
<td>32.95%</td>
<td>4.06%</td>
<td>44.73%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liquid gels</td>
<td>$204.55</td>
<td>$140.66</td>
<td>$84.88</td>
<td>$14.43</td>
<td>$5.51</td>
<td>$21.58</td>
<td>0</td>
<td>$23.06</td>
<td>$44.74</td>
<td>0</td>
<td>0</td>
<td>$206.82</td>
<td></td>
</tr>
<tr>
<td></td>
<td>33.34%</td>
<td>27.57%</td>
<td>22.51%</td>
<td>4.77%</td>
<td>1.88%</td>
<td>7.95%</td>
<td>0</td>
<td>11.26%</td>
<td>38.43%</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Legs/muscle</td>
<td>$101.42</td>
<td>$133.31</td>
<td>$17.53</td>
<td>$64.17</td>
<td>$29.96</td>
<td>$18.38</td>
<td>$23.88</td>
<td>$23.06</td>
<td>$56.44</td>
<td>$7.54</td>
<td>$204.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>16.53%</td>
<td>26.13%</td>
<td>4.65%</td>
<td>21.21%</td>
<td>10.22%</td>
<td>6.77%</td>
<td>9.59%</td>
<td>11.15%</td>
<td>19.81%</td>
<td>48.66%</td>
<td>6.65%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gentle on stomach</td>
<td>$66.32</td>
<td>$45.46</td>
<td>$24.40</td>
<td>$23.18</td>
<td>$55.50</td>
<td>$55.21</td>
<td>$38.35</td>
<td>$44.74</td>
<td>$23.06</td>
<td>$0.88</td>
<td>0</td>
<td>$116.40</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10.81%</td>
<td>8.91%</td>
<td>6.47%</td>
<td>7.66%</td>
<td>18.93%</td>
<td>20.34%</td>
<td>15.40%</td>
<td>21.63%</td>
<td>11.26%</td>
<td>0.76%</td>
<td>14.83%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Back</td>
<td>$25.65</td>
<td>$51.94</td>
<td>$15.88</td>
<td>$44.99</td>
<td>$39.29</td>
<td>$19.22</td>
<td>$4.71</td>
<td>0</td>
<td>$56.44</td>
<td>$0.88</td>
<td>$18.83%</td>
<td>$115.98</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.18%</td>
<td>10.18%</td>
<td>4.21%</td>
<td>14.87%</td>
<td>13.40%</td>
<td>7.08%</td>
<td>1.89%</td>
<td>0</td>
<td>27.56%</td>
<td>0.76%</td>
<td>13.10%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fewer pills</td>
<td>0</td>
<td>$53.01</td>
<td>$6.22</td>
<td>$100.06</td>
<td>$30.46</td>
<td>$80.12</td>
<td>$50.65</td>
<td>0</td>
<td>$7.54</td>
<td>0</td>
<td>$14.83</td>
<td>$113.25</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>10.39%</td>
<td>1.65%</td>
<td>33.07%</td>
<td>10.39%</td>
<td>29.52%</td>
<td>20.34%</td>
<td>0</td>
<td>3.68%</td>
<td>0</td>
<td>12.79%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

\(^6\) The remaining 7 attributes (not reported) had negligible advertising expenditures. The sum of the expenditures in Figure 3 exceeds the total ad spending because many advertisements promote multiple characteristics, and for the purpose of this Figure only, we attributed total ad spending to each characteristic mentioned.
We examine the attribute usage correlations to investigate whether the coded cues represent distinct information. For example, although we code “strong” and “fast” as separate information categories, we also ensure that the coded information categories are indeed distinct information cues. Table 2 portrays the correlation matrix of cue usage and shows that the cue descriptors we use are distinctive. For example, both “fast” and “strong” are often used together, but in more than half the occurrences (in dollar terms), they are also used separately. Thus, two cues may be used together frequently, but each still provides important information to consumers. “Strong” denotes how powerful the medicine is, and “fast” denotes speed of the onset of pain relief.

Identification Strategy

A brand’s decision about how much information to include in an advertisement is likely to be made simultaneously with the decision about the type of advertisement (comparative or noncomparative). Therefore, these two decisions are interdependent, much in the same way as equilibrium price and quantity are determined at the same time in a simple demand–supply model. In other words, there is some unobservable exogenous variable that explains both the information content and whether that content is a noncomparative or comparative advertisement. For example, a higher-quality brand might provide less information and have more comparative advertisements than a lower-quality brand.

The other two potentially endogenous explanatory variables we consider are the size of the brand and the size of a brand’s generic counterpart. To see why endogeneity might be an issue, note that market size and information content are outcomes of brands’ strategic interactions, and a fully structural equilibrium model would specify three equations, one for each

---

7 There are two instances of high correlation that merit comment. First, whenever "liquid gels" are mentioned, "fast" is almost always mentioned. Second, "long lasting" and "fewer pills" are often used together, or 33% of the time. Conversely, when "fewer pills" are mentioned, long lasting was mentioned 88% of the time. In this instance, we could provide an umbrella classification that encompasses both, but the difference in results would be minor.
of the three variables (information content, market size, and generic counterpart market size).

Here, we estimate only one of the three equations—the one that explains information content as a function of the other two endogenous variables—but we control for the endogeneity of the other two variables. Equivalently, we could consider an unobservable variable (e.g., quality) that is correlated with the brand’s market size and information content. By omitting that variable from our regressions, we would introduce a bias in the estimation of the parameters of the model.

We use an instrumental variable approach to determine whether our concerns about the endogenous variables are empirically relevant (Villas-Boas and Winer 1999). Instruments that correlate with sales and advertising but not with unobserved quality provide information on how important the endogeneity problem is likely to be. Following the literature on the estimation of demand in differentiated product markets (e.g., Levinsohn, and Pakes 1995; Nevo 2001; Chintagunta 2001), we assume that the product characteristics space is exogenous. This is a particularly reasonable assumption for the OTC analgesics industry because pain relievers’ true characteristics are essentially fixed, determined by the chemical properties of the particular active ingredient that constitutes the drug and regulated by the FDA. Then, we consider the case when a brand’s own characteristics enter directly in the ordered probit regression, and we construct the instrumental variables using functions of the characteristics of the brand’s competitors. More specifically, we construct both the means of the characteristics of the brand’s competitors and their minimum values. We also interact these characteristics with a dummy that is equal to 1 if a brand’s parent company owns another brand, which is the case for Tylenol and Motrin (parent company McNeil) and Aleve and Bayer (parent company Bayer). Finally, we
interact the characteristics with the 2005 year dummy to capture advertising content changes; 2005 was considered one of the most turbulent years in the analgesics industry.\textsuperscript{8}

To deal with the endogenous variables in our nonlinear ordered probit model, we follow Rivers and Vuong’s (1988) proposed approach. First, we rewrite the information content as

\[
\alpha_t - X\beta - w\gamma = \bar{p} - \bar{I} - \frac{\bar{I}}{\bar{S}},
\]

where \(w\) is a vector of the three endogenous variables. The main identification assumption is that instruments are not correlated with the error term (i.e., \((Z'\varepsilon) = 0\)). Here \(Z\) includes all the exogenous variables, such as brand characteristics \((X)\) and functions (here, the average) of the characteristics of the brand’s competitors. We use a two-step procedure described in Rivers and Vuong (1988). First, we run an ordinary least squares regression,

\[
w = Z\delta + v,
\]

where \(v\) is not observed; this is the omitted variable that generates the endogeneity problem. This first stage regression yields residuals \(\hat{v}\), which we include in the ordered probit in the second stage of the estimation. The estimation of this ordered probit provides consistent estimates of all the parameters. We explain the approach with multiple endogenous regressors in greater detail in Appendix B. Here, we are not interested in the magnitude of the parameters of the information content relationship (i.e., \(\beta\)) but rather in the marginal effects of a change in the endogenous variables, which we can consistently estimate using Blundell and Powell’s (2003) approach.

**Empirical Findings**

Table 3 reports the findings of the various specifications estimated with our data. The first five columns present the results from estimating our model with the attribute coding. For robustness, in the last column we also include results estimated with the RS information coding.

\textsuperscript{8} The growth of information content during 2005 is most likely due to the FDA’s announcement at the end of 2004 of the results of a clinical study, which indicated that patients taking naproxen sodium (Aleve) may be at an increased risk of suffering heart attack or stroke (the withdrawal of Vioxx was also associated with this clinical study). By the end of January 2005, sales of Aleve plummeted by more than 50%, suffering the largest decline in brand history (for more details, see ARF Ogilvy Awards 2007).
**H1: Vertical Characteristics**

We test H1, i.e. that brands with higher vertical quality transmit more information, by associating the number of cues, which is our dependent variable, with the values of the exogenous medical characteristics of the active ingredients. The first column of Table 3 includes NNT, relative speed, and measures of cardiovascular and gastrointestinal risk as such explanatory variables.

Keeping all else constant, we find that brands with inherently greater strength of pain relief have advertisements with more information content. Recall that for NNT, a higher number means worse performance and a less effective drug. Thus, the negative coefficient on NNT is consistent with strong and efficient drugs having more information in their advertisements. We find a similar pattern for brands that have lower gastrointestinal and cardiovascular risks: Their advertisements also tend to be more informative. Finally, brands that offer faster relief also have more information content.

Overall, the first column of Table 3 suggests that more information is provided the stronger is a brand in one of the four dimensions identified by the exogenous medical characteristics of the active ingredient. This finding provides empirical support for H1.
Table 3. Determinants of Information Disclosure

<table>
<thead>
<tr>
<th></th>
<th>Attribute Coding</th>
<th>R-S Coding</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Comparative?</td>
<td>0.559***</td>
<td>2.115***</td>
</tr>
<tr>
<td></td>
<td>(0.025)</td>
<td>(0.099)</td>
</tr>
<tr>
<td>Standardized Sales</td>
<td>0.445***</td>
<td>0.268***</td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
<td>(0.047)</td>
</tr>
<tr>
<td>Standardized Sales Squared</td>
<td>-0.204***</td>
<td>-0.160***</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.025)</td>
</tr>
<tr>
<td>Standardized Generic Sales</td>
<td>-0.136***</td>
<td>-1.571***</td>
</tr>
<tr>
<td></td>
<td>(0.043)</td>
<td>(0.084)</td>
</tr>
<tr>
<td>Standardized NNT</td>
<td>-0.724***</td>
<td>-0.602***</td>
</tr>
<tr>
<td></td>
<td>(0.029)</td>
<td>(0.038)</td>
</tr>
<tr>
<td>Relative Speed</td>
<td>0.336***</td>
<td>0.281***</td>
</tr>
<tr>
<td></td>
<td>(0.041)</td>
<td>(0.042)</td>
</tr>
<tr>
<td>Standardized GI Risk</td>
<td>-0.133***</td>
<td>-0.132***</td>
</tr>
<tr>
<td></td>
<td>(0.035)</td>
<td>(0.035)</td>
</tr>
<tr>
<td>Standardized CV Risk</td>
<td>-0.151***</td>
<td>-0.185***</td>
</tr>
<tr>
<td></td>
<td>(0.029)</td>
<td>(0.030)</td>
</tr>
<tr>
<td>Residuals-Comparative</td>
<td>-1.652***</td>
<td>-1.103***</td>
</tr>
<tr>
<td></td>
<td>(0.102)</td>
<td>(0.178)</td>
</tr>
<tr>
<td>Residuals-Sales</td>
<td>0.174</td>
<td>1.056***</td>
</tr>
<tr>
<td></td>
<td>(0.117)</td>
<td>(0.124)</td>
</tr>
<tr>
<td>Residuals-Generic Sales</td>
<td>1.844***</td>
<td>0.512***</td>
</tr>
<tr>
<td></td>
<td>(0.112)</td>
<td>(0.117)</td>
</tr>
<tr>
<td>Cutoff (0-&gt;1 Cues)</td>
<td>-3.085***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.194)</td>
<td></td>
</tr>
<tr>
<td>Cutoff (1-&gt;2 Cues)</td>
<td>-1.041***</td>
<td>-0.941***</td>
</tr>
<tr>
<td></td>
<td>(0.090)</td>
<td>(0.094)</td>
</tr>
<tr>
<td>Cutoff (2-&gt;3 Cues)</td>
<td>0.291***</td>
<td>0.436***</td>
</tr>
<tr>
<td></td>
<td>(0.089)</td>
<td>(0.093)</td>
</tr>
<tr>
<td>Cutoff (3-&gt;4 Cues)</td>
<td>1.095***</td>
<td>1.271***</td>
</tr>
<tr>
<td></td>
<td>(0.090)</td>
<td>(0.094)</td>
</tr>
<tr>
<td>Cutoff (4-&gt;5 Cues)</td>
<td>2.032***</td>
<td>2.228***</td>
</tr>
<tr>
<td></td>
<td>(0.091)</td>
<td>(0.096)</td>
</tr>
<tr>
<td>Cutoff (5-&gt;6 Cues)</td>
<td>3.034***</td>
<td>3.234***</td>
</tr>
<tr>
<td></td>
<td>(0.097)</td>
<td>(0.101)</td>
</tr>
<tr>
<td># of Obs.</td>
<td>9,739</td>
<td>9,708</td>
</tr>
<tr>
<td>Log-Likelihood</td>
<td>-13847.5</td>
<td>-13527.6</td>
</tr>
</tbody>
</table>

Note: *** p<0.01, ** p<0.05, * p<0.1; Bootstrapped standard errors reported in Columns 2, 5 and 6. Endogenous variables: Comparative?, Standardized Sales, Standardized Generic Sales. Instruments: (1) averages of: Standardized GI Risk, Standardized CV Risk, Standardized NNT, Relative Speed; (2) Minimums of: Standardized GI Risk, Standardized CV Risk; (3) Interactions between characteristics and year 2005 dummy; (4) Interactions between characteristics and a dummy indicating whether a brand has a parent company that owns competing brand in the category. The first stage R² for endogenous variables is the following: Comparative? - R² = 0.28, Standardized Sales - R² = 0.95, Standardized Generic Sales - R² = 0.93)

As we discussed in our theoretical section, the cutoffs estimated in these regressions have a clear structural interpretation. In particular, we showed that $\alpha_{t+1} - \alpha_t = \frac{L - L_{t+1}}{3}$. In our analysis
\( \bar{s} \) is unknown, so assume for interpretation that \( \bar{s} = 1.9 \). Then, for each \( i \) we can compute the difference \( I_i - I_{i+1} \), which is the incremental benefit of the \( i^{th} \) information cue \( i \). Consider the case of \( i=5 \). Then, \( I_5 - I_6 = 1.002 \). We can then find the point \((6, 1.002)\) in Figure 3 to represent the empirical analogue of the theoretical incremental benefit corresponding to five information cues that we had derived in Panel B of Figure 2. We can replicate this exercise for the case of \( i=4 \), for which the scatter point will correspond to \((5, 1.939)\), where is equal to \( 1.002 + 0.937 \). By repeating this exercise for all \( i =1,…,6 \), we derive the scatter plot in Figure 3. What we see in Figure 3 is remarkably similar to what we derive theoretically in Panel B of Figure 2, which provides empirical support for our theoretical model: there is a fundamental trade-off between information and persuasion content that firms face when preparing an advertisement. This relationship, depicting the diminishing marginal returns to information, holds in every specification that we estimate and looks very similar to the one illustrated in Figure 3.

---

Instead of normalizing \( \bar{s} = 1 \), one could also report \( \frac{a_{i+1} - a_i}{a_6 - a_1} \), but that would only rescale Figure 3. The normalization \( \bar{s} = 1 \) gives a more intuitive interpretation for Figure 3.
**H2: Comparative Versus Noncomparative Advertising**

According to H2, brands include more information in comparative than noncomparative advertisements. Note that comparative advertisements will always have at least one cue because a brand must compare itself with another brand on at least one dimension. Therefore, the analysis examines whether comparative and noncomparative ad information content is different, conditional on having at least one information cue.

The second and third columns in Table 3 present ordered probit results of advertising content as a function of whether an advertisement was comparative or not. The second column treats the choice of comparative advertisement as an exogenous variable. The third column treats the choice as an endogenous variable. The second column shows that the comparative advertising dummy is highly statistically significant, as indicated by the positive value of the coefficient. This result indicates that comparative advertisements have more informational cues and that the likelihood that an advertisement is comparative increases with the number of cues.

The third column allows for endogeneity in the comparative advertising decision. We find that the parameter estimate for the “Comparative” dummy is much higher, implying that comparative advertisements have more information content than noncomparative advertisements and that such information content is significantly higher than under the exogenous treatment. The strong endogeneity of this dummy variable is confirmed by the estimated coefficient of the control function, which is statistically significant at the 1% confidence level. Thus, we cannot reject the hypothesis that the variable “Comparative” is endogenous (Smith and Blundell 1986). Therefore, our omitted variable concern in this setting is valid, and such unobserved quality is associated with the attractiveness of having comparative advertisements. This result provides strong support for H2.
We compute the marginal effects for comparative advertising using the results in the second and third columns of Table 3. The marginal effects appear in Table 4. In particular, under the exogenous treatment, we show that when an advertisement has comparative content, it is 14.4% less likely to include only one cue and 13.9% more likely to include five cues. Table 4 also shows the marginal effects for endogenous specification. Under this treatment, the probability of an advertisement containing only one cue is 33.7% less likely in comparative than noncomparative advertisements. Similarly, we find that the probability of observing five cues is 31.2% more likely in comparative than noncomparative advertisements. These marginal effect differences between endogenous and exogenous treatments suggest that the bias introduced by the endogeneity of the comparative advertising choice is very large.

<table>
<thead>
<tr>
<th>Table 4. Marginal Effects of Decision to do Comparative Ads</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marginal effects (by number of cues)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>1   2   3   4   5   6</td>
</tr>
<tr>
<td>Exogenous Treatment</td>
</tr>
<tr>
<td>Comparative?</td>
</tr>
<tr>
<td>-0.144***       -0.188***       0.007       0.134***       0.139***       0.047**</td>
</tr>
<tr>
<td>(0.032)         (0.013)         (0.095)       (0.021)         (0.040)         (0.023)</td>
</tr>
<tr>
<td>Endogenous Treatment</td>
</tr>
<tr>
<td>Comparative?</td>
</tr>
<tr>
<td>-0.337***       -0.434***       0.024       0.256**       0.312**       0.165**</td>
</tr>
<tr>
<td>(0.070)         (0.085)         (0.243)       (0.121)         (0.132)         (0.082)</td>
</tr>
</tbody>
</table>

**H3 and H4: Brand Market Size and Generic Counterpart Market Size**

To test H3 and H4, we include the variables that measure brand size and brand size squared (to capture a possible nonlinear relationship) and the size of the brand’s generic counterpart. Table 3 presents the results of the ordered probit with standardized sales, squared standardized sales, and standardized generic sales. In this specification, we conclude that the largest brands transmit less information compared to the rest of the brands. There are very few observations for medium sized brands, therefore it is difficult to conclude whether the negative
size-information content relationship holds over the entire range of brand sizes (the positive
coefficient on brand size variable and the negative coefficient on the quadratic term suggests that
the relationship resembles a weak inverse U-shape).

The evidence is also consistent with a relatively large spillover effect from informative
advertising: We find that branded firms include less information content when the size of their
generic competitors is large because the parameter of standardized generic sales is negative.

We also find that the coefficient estimate for the comparative advertisement dummy is
close to the one in the second column of Table 3. Therefore, we conclude that comparative
advertising decisions are not collinear with brand sales and generic sales. This result is
noteworthy because it suggests that the comparative advertisement decision might not depend
significantly on the market share of the attacking brand but rather on the market share of the
target brand (i.e., attack larger brands). Anderson et al. (2010) investigate the relationship among
the attacker’s market share, the attacked brand’s market share, and the amount of their
comparative advertising. They show that the decision to use comparative advertising is due to the
market share of the attacked brand and the interaction between the shares of the attacked brand
and those of the attacking brand.

In the fifth column of Table 3 we follow an instrumental variable approach to estimate
the effect of comparative advertisements, sales, and generic sales on information content. The
estimated coefficient of comparative advertisements is large, though not as much as it is in the
third column. The coefficient on standardized sales is similar to the coefficient under exogenous
treatment; however, the marginal impact of the generic counterpart’s market share is
significantly greater under endogenous treatment but continues to be negatively related to
information content.
This result implies that when endogeneity of the generic counterpart market share is not accounted for, the negative relationship between it and advertising content is significantly underestimated. In line with the exogenous treatment in the fourth column, we observe the same relationship between the brand size and information content. Thus, these results indicate that the largest brands include less information content.

We also observe that the coefficient estimates of the control functions for the generic sales (1.844) and for the decision to have comparative advertising (-1.103) are both statistically significant, proving that both variables (generic sales and comparative advertisements) are endogenous. The estimated coefficient of the control function for the brand sales (0.174) is only statistically significant at the 13.7% level. This finding is probably due to our instrumental variables having just enough identification power to identify two endogenous variables.

To understand the economic importance of the results regarding brand market size, we constructed figures that associate a brand’s probability of choosing a given amount of information (e.g., one cue) with the distribution of brand size (at the 10th, 25th, 50th, 75th, and 90th percentiles of market size distribution). For example, Figure 6 shows that the likelihood of an advertisement including just one cue increases sharply with size. In contrast, the probability of observing an advertisement with four or more cues decreases with size. For example, a move from the 25th percentile in the size distribution to the 75th percentile increases the probability of providing only one cue by approximately 25%.
Similarly, Figure 7 represents the changes in the likelihood that an advertisement includes a certain number of information cues depending on the size of generic counterpart market share (again, evaluated at the 10th, 25th, 50th, 75th, and 90th percentiles of generic counterpart market size distribution). For example, the likelihood that a product with a large generic counterpart will include one information cue is 80% higher than the likelihood that a small generic counterpart product will include one information cue. The converse is true for
three or more information cues: Smaller generic counterpart brands are significantly more likely to include more information in their advertisements.

In summary, we find weak support for H₃ and strong support for H₄.

Figure 7. H₄: Marginal effects on Information Disclosure by Generic Counterpart Size
**RS Approach**

The last column of Table 3 presents the results of the estimation using the RS information category coding approach. We implement this step to evaluate the extent to which our results are robust to alternative information content definitions. We find that all the hypotheses (except for H1) are also supported with the RS coding approach. All the key variables (except for vertical quality) have the same signs as those with the attribute coding approach. The magnitudes are smaller because the left-hand-side variable (information content) under the RS coding approach is smaller (see Figure A2 in the Appendix C).

The vertical characteristics parameter estimates under the RS information coding approach are different for strength (NNT) and speed. As we explained previously with regard to this mismatch, under the RS approach, both characteristics fall under one information category and therefore understate the information content of advertising.

**Final Remarks**

This research provides several contributions to the literature on information content of advertising. First, we develop a theoretical model that describes the tradeoff between persuasive and informative content within an advertisement. We use the theoretical model to predict the relationship between the informative content and various market performance indicators. The empirical component of the paper applies the theory to the OTC analgesics advertising content data. The analysis gives rise to a deeper and broader description of information content, which combines advertising content knowledge with industry structure and advertising spending data.

In line with our theoretical results of a trade-off between informative and persuasive advertising, we show empirical evidence that supports our hypotheses: we find that brands with higher vertical quality disclose more information and high market share brand advertisements
provide less information. These two results are not contradictory: The endogeneity of brand size underscores the importance of correcting for it; otherwise, it might be assumed that larger brands are fundamentally of higher quality than smaller brands and, therefore, that their advertisements should have more information content. We also show that, after correcting for endogeneity, more competition from generics gives rise to less information transmitted by branded products, which is in line with the view that there are significant spillovers to informative advertising. Finally, we quantify the extent to which comparative advertisements have significantly more information content than noncomparative ones; the effect is much larger than would be predicted without correcting for endogeneity.

From a methodological perspective, we describe a method that is appropriate for dealing with information content with multiple explanatory variables; and show how the analysis can be corrected for endogeneity and how this alters the results.

Our empirical analysis is restrictive in several aspects and therefore suggests extensions that constitute themes for further research into information content. First, further research could use a subclassification of cues (e.g., into vertical cues, such as “fast,” that all consumers appreciate or horizontal cues, such as “headache” or “menstrual pain” that only some consumers desire) to explore the differential content of different cue types. Second, an information cue can be deployed only if a product has the attribute communicated in the cue and can be used comparatively only if a product has an advantage over another product. Thus, further research might examine the amount of information advertised as a function of the total number of cues that could feasibly be advertised. Likewise, further research might explore whether products use comparative advertisements more against similar or dissimilar products. Third, we do not examine product advertising campaigns, in which advertisements address a subset of themes over a limited horizon. Fourth, we code only the objective content of advertisements as
quantified by their reference of specific characteristics and competitors. We recognize that advertising may persuade through channels other than pure information and lead consumers to act on emotional factors. However, we have not attempted to code such effects, but our theory model allows for such persuasive effects to exist. The primary purpose of the empirical component of this paper, was to measure objective content of advertising along the lines of traditional content analysis. Incorporating the subjective side would be an important aspect to explore in further extensions. Finally, we do not address whether market provision of information is optimal or how valuable the information is to consumers (Ippolito and Pappalardo 2002; Pappalardo and Ringold 2000). The purpose here was to document and, using our theoretical model, to rationalize empirical regularities present in the data and to provide measures of the fundamental variables that can be used to answer such questions.

The current theory model and content analysis methodology can readily be applied to other product categories and industries. Extensions of the proposed methodology could be applied to answer questions such as the following: How does advertising information content differ for experience, search, and credence products? Does the relationship between brand size and information hold in more broad contexts? Do new products provide more information? These questions should hold both empirical and theoretical interest.
REFERENCES


Appendix A: Explanation of Medical Measures

The medical literature provides objective risk and efficiency measures for each product, based on its active ingredient (or combination of ingredients), strength, and recommended dosage. Each active ingredient has definitive maximum doses and durations of therapy. Differences exist across different active ingredients in terms of the important safety issue of the potential for gastrointestinal toxicity and cardiovascular risk. We collected the measurable characteristics for maximum OTC recommended dosage (single dose): Ibuprofen: 400 mg; naproxen sodium: 440 mg; aspirin: 1000 mg; and acetaminophen: –1000 mg. These measurements enable us to identify the locations of all the products in the characteristics space.

Relative risk is the risk of an event (e.g., developing a disease) relative to exposure. Relative risk is the ratio of the probability of the event (E) occurring in the exposed group versus the control (nonexposed) group:

\[
RR = \frac{Pr(E|treatment)}{Pr(R|control)}
\]

Relative risk is used frequently in clinical trial data to compare the risk of developing a disease in people not receiving the new medical treatment (or receiving a placebo) versus people receiving an established (standard of care) treatment. In the case of the gastrointestinal and cardiovascular relative risk numbers used herein, we use them to compare the risk of developing a side effect in people receiving a drug with people who do not receive the treatment (or receive a placebo). Thus, a cardiovascular relative risk of 1.44 means that cardiovascular problems arise with 44% higher likelihood using the drug (versus placebo).

We compute NNT with respect to two treatments, A and B, with A typically a drug and B a placebo. If the probabilities \( P_A \) and \( P_B \) under treatments A and B, respectively, are known, we can compute NNT as follows:

\[
NNT = \frac{1}{P_B - P_A}
\]

The NNT for a given therapy is simply the reciprocal of the absolute risk reduction (ARR= \( P_B - P_A \)) for that treatment. For example, in a hypothetical migraine study, if risk decreased from \( P_B = .30 \) without treatment with drug M to \( P_A = .05 \) with treatment with drug M, for a relative risk of .17 (\( .30 / .05 \)), a relative risk reduction of .83 (\( .3 - .05 \)/.3), and an absolute risk reduction of .25 (.3 – .05), the NNT would be 1/.25, or 4. In clinical terms, an NNT of 4
means that four patients need to be treated with drug M to prevent migraine from recurring in one patient. Typically, the lower the NNT number, the more potent and efficient the treatment is.

Table A1. Relative Rankings of Speed and Duration of Pain Relief

<table>
<thead>
<tr>
<th>Time to Perceptible Pain Relief</th>
<th>Duration of Meaningful Time Relief (Longevity)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sol Ibuprofen&gt; Ibuprofen (1, 6)</td>
<td>Naproxen&gt; Aspirin (3)</td>
</tr>
<tr>
<td>Ibuprofen&gt; Acetaminophen (1, 5, 6)</td>
<td>Ibuprofen&gt; Acetaminophen (2, 4, 5, 9, 10)</td>
</tr>
<tr>
<td>Acetaminophen&gt; Naproxen (10)</td>
<td>Ibuprofen/Sol ib&gt; Acetaminophen (6)</td>
</tr>
<tr>
<td>Naproxen&gt; Aspirin (3)</td>
<td>Naproxen&gt; Acetaminophen (2, 4, 8, 9)</td>
</tr>
<tr>
<td>Acetaminophen&gt; Naproxen Sodium (10)</td>
<td>Acetaminophen&gt; Aspirin (10)</td>
</tr>
</tbody>
</table>

Resulting Ranking (Highest to Lowest):

1. Soluble Ibuprofen
2. Ibuprofen
3. Acetaminophen
4. Naproxen Sodium
5. Aspirin

We reviewed several medical journal articles to rank the three efficiency measures (maximum level of pain relief achieved, onset to perceptible pain relief, and duration of meaningful pain relief) of the analyzed active ingredients. Most medical articles compare only two or three active ingredients. If article X noted that drug A is more efficient than drug B (A > B) and article Y noted that drug B is more efficient than C (B > C), we conclude by assuming that A is more efficient than B and C (A > B > C). Subsequently, we also present the numbered list of references we used to infer relative rankings. Table A1 lists all the relative relationships and references of medical articles and gives the resulting ranking presented in Figure 2.


(5) Cooper, S.A., B.P. Schachtel, E. Goldman, S. Gelb, and P. Cohn (1989), "Ibuprofen and


**Appendix B. Two-Step Procedure to Correct for Endogeneity**

To deal with endogenous variables in our nonlinear ordered probit model, we follow Rivers and Vuong’s (1988) approach. First, we rewrite the information content as $X\beta + w\gamma + \epsilon$, where $w$ (e.g., market share) is an endogenous variable: $E(w'\epsilon \neq 0)$. The main identification assumption is that there is a matrix of variables $Z$ (which contains $X$), such that $E(Z'\epsilon = 0)$. Here, $Z$ includes all the exogenous variables, such as $X$ and functions of the characteristics of the brand's competitors. Let $w = Z\delta + \nu$, which indicates how the variation in the exogenous variables $Z$ explains the variation in the endogenous variable $w$. We assume that $(\nu, \epsilon, Z)$ are i.i.d., and $\nu$ and $\epsilon$ have, conditional on $Z$, a joint normal distribution:

$$\begin{pmatrix} \epsilon \\ \nu \end{pmatrix} \sim N\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{\epsilon\nu}\sigma_\epsilon\sigma_\nu \\ \rho_{\epsilon\nu}\sigma_\epsilon\sigma_\nu & \sigma_\epsilon^2 \end{pmatrix} \right)$$

We can then use a basic property of jointly normal random variables and write:

$$E(\epsilon|\nu) = \frac{\rho_{\epsilon\nu}}{\sigma_\nu} \sigma_\epsilon \nu = \frac{\rho_{\epsilon\nu}}{\sigma_\nu} \nu$$

Let the error $\epsilon$ be defined as follows:
\[ \varepsilon = \varepsilon - E(\varepsilon | \nu) = \varepsilon - \frac{\rho_{ev}}{\sigma_v} \nu. \]

Then, \( \varepsilon \sim N(0,1 - \rho_{ev}^2) \), and then the information content is \( X\beta + w\gamma + \frac{\rho_{ev}}{\sigma_v} \nu + \varepsilon. \)

If we observe \( \nu \), we can again run a standard ordered probit model because the inclusion of \( \nu \) controls for the endogeneity of \( w \). Here, we would only need to pay extra attention when estimating the parameters. Then, we would have the following:

\[
\Pr(y = j | X) = \Phi(\alpha_{j\rho} - X\beta_{\rho} - w\gamma_{\rho} - \theta_{\rho} \nu) - \Phi(\alpha_{j-1,\rho} - X\beta_{\rho} - w\gamma_{\rho} - \theta_{\rho} \nu),
\]

where each parameter has been rescaled:

\[
\alpha_{j\rho} = \frac{\alpha_j}{\sqrt{1 - \rho_{ev}^2}}, \beta_{\rho} = \frac{\beta}{\sqrt{1 - \rho_{ev}^2}}, \gamma_{\rho} = \frac{\gamma}{\sqrt{1 - \rho_{ev}^2}} \text{ and } \theta_{\rho} = \frac{\rho_{ev}}{\sigma_v}.
\]

The crucial observation here is that we are not interested in the parameters of the information content relationship (e.g., \( \beta \)'s) but rather in the marginal effects of a change in the endogenous variables. Then, following Blundell and Powell (2003), we can consistently estimate the average structural function as follow

\[
A\bar{S}F(X, w) = N^{-1} \sum_{i=1}^{N} \Phi \left( \hat{\alpha}_{j\rho} - X_i\hat{\beta}_{\rho} - w_i\hat{\gamma}_{\rho} - \hat{\theta}_{\rho} \hat{\nu}_i \right) - \Phi \left( \hat{\alpha}_{j-1,\rho} - X_i\hat{\beta}_{\rho} - w_i\hat{\gamma}_{\rho} - \hat{\theta}_{\rho} \hat{\nu}_i \right)
\]

which we then use to rescale the coefficient of the variable the marginal effect of which is of interest. For example, to obtain the marginal effect of a change in \( w \), we multiply \( A\bar{S}F(X, w) \) by \( \hat{\gamma}_{\rho} \).

**Appendix C. Content Analysis**

**Expenditure-Weighing**

Many traditional content studies do not include data on what brands paid to screen, air, or publish advertisements, so only information cues per advertisement analyzed are reported. As we discussed previously, our data include the complete set of advertisements run over the sample period and the expenditure and airing frequency data. These data provide different ways to count the observations. Figure A1 represents the histograms of information cue distributions under varying approaches. The first columns in Figure A1 weight each advertisement occurrence by its advertising expenditures. The "unweighted" data in Figure A1 use each separate advertisement within any given month as an observation. In this case, multiple airings of the same advertisement within the same month are ignored. This is in line with most of the traditional
content analysis studies: Multiple copies of the same advertisement were typically not counted as
different observations. The middle columns weight each advertisement by airing frequency.
Finally, the last columns treat each unique advertisement as a different observation. Arguably,
the first approach (weighting by advertising expenditures) is the best measure of the economic
importance of each information category and the extent of information content relayed to
consumers. However, the distribution of information cues across the different measures is
similar, which indicates that a lack of data on the amount spent on airing the advertisements does
not distort the results for our particular sample. The distribution of cues for the unweighted data
stochastically first-order dominates the distribution for the weighted data -- there is a greater
fraction of ads with zero cues, a greater fraction with one or fewer, etc. This implies that more
money tends to be spent on running ads with more cues in them. As a benchmark, if all ads cost
the same amount to screen, then ads with more cues are being screened more often.
Alternatively, if they were all aired the same number of times, those with more cues are being
aired to more expensive (i.e., typically larger) audiences.

Figure A1. Distribution of Number of Cues

![Distribution of Number of Cues](image)

Comparing RS and attribute coding approaches.

Figure A2 shows the cue distribution and descriptive statistics for comparative and
noncomparative advertisements under the attribute and RS coding approaches. We find that
under the attribute approach, comparative advertisements have on average (weighted by
expenditure) 3.36 information cues, whereas noncomparative advertisements have 2.61 cues.
Under the RS information classification approach, these numbers are 2.25 and 2.12, respectively.
These findings are broadly consistent with the findings of Harmon, Razzouk, and Stern (1983) and Chou, Franke, and Wilcox (1987) for magazine advertisements. Harmon, Razzouk, and Stern find that comparative advertisements have 1.84 cues and noncomparative advertisements have only .86 cues. In our sample, on average, each advertisement contains more information, and the difference between the number of cues for comparative and noncomparative advertisements is much smaller.

Figure A2. Distribution of Number of Cues: Attribute Coding vs. Resnik-Stern (RS)