Correcting Audience Externalities in Television Advertising

Kenneth C. Wilbur, Linli Xu, David Kempe

Please scroll down for article—it is on subsequent pages

INFORMS is the largest professional society in the world for professionals in the fields of operations research, management
science, and analytics.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org

Correcting Audience Externalities in Television Advertising

Kenneth C. Wilbur
Rady School of Management, University of California, San Diego, California 92093, ken@kennethcwilbur.com

Linli Xu
Carlson School of Management, University of Minnesota, Minneapolis, Minnesota 55455, linlixu@umn.edu

David Kempe
Department of Computer Science, University of Southern California, Los Angeles, California 90033, dkempe@usc.edu

When a television advertisement causes viewers to switch channels, it reduces the audience available to subsequent advertisers. This audience loss is not reflected in the advertisement price, resulting in an audience externality. The present article analyzes the television network's problem of how to select, order, and price advertisements in a break of endogenous length in order to correct audience externalities. It proposes the Audience Value Maximization Algorithm (AVMA), which considers many possible advertisement orderings within a dynamic programming framework with a strategy-proof pricing mechanism. Two data sets are used to estimate heterogeneity in viewer-switching probabilities and advertiser willingness-to-pay parameters in order to evaluate the algorithm's performance. A series of simulations shows that AVMA typically maximizes audience value to advertisers, increases network revenue relative to several alternatives, and runs quickly enough to implement.

1. Introduction

Television networks sell commercials to the highest-bidding advertisers based on expected program audiences. Then, the networks order ads randomly within the break. Advertising prices do not depend on ad content or on the audience loss a particular ad may cause, even though that audience loss reduces the number of viewers who remain to watch the rest of the commercial break. Strategic advertisers therefore may have an incentive to include ad stimuli that increase their own effectiveness at the expense of the network's remaining audience size. This unpriced divergence of interests is called an *audience externality*.

Historically, audience externalities have not been observable. In the United States, sales of national television ads are based on audience estimates from Nielsen's National Television Index (NTI) sample. This sample includes approximately 21,000 households, less than 0.02% of the population (Gaither 2011). Nielsen's sampling errors are too large to allow for meaningful comparisons between consecutive advertisements' audiences.

The precise measurement of commercial audiences has recently become possible. In 2012, 83.5% of U.S. households received television signals through digital cable or direct-broadcast satellite systems (TVB 2013).[1] The set-top boxes that connect televisions to these digital signal distribution networks can passively record and report complete moment-by-moment tuning data for each household. Thus, audience *estimates* can now be replaced with audience *measurements*.

The present article considers how a network could use commercial-specific audience measurements to correct audience externalities in television advertising. Adopting such an approach would require a change in business model, which may be costly. We study what such a change might look like and when it might be profitable.

The television network faces a complex problem, even for a single commercial break: given J slots and n potential advertisements, the selection/ordering problem alone has $\binom{n}{J} J!$ possible solutions. We offer a model of audience retention and value, and then we develop algorithms to determine advertisement selection, ordering, and pricing within a commercial break of endogenous length, accounting for slot-specific tune-away rates and anticompetition constraints. The paper proposes the Audience Value Maximization Algorithm (AVMA), a solution that selects, orders, and

[1] This figure has doubled since 2004.

prices ads to maximize audience value to advertisers and then chooses the break length to maximize network profits.

Market data are employed to learn about viewer switching and to assess the performance of the proposed algorithms. Raw data and model parameter estimates reveal substantial heterogeneity in tune-away rates, both across advertising creatives and across viewing segments. Ignoring this heterogeneity leads to biased inferences and missed opportunities to optimize viewer retention. A series of simulations are conducted based on the parameter estimates, explicitly accounting for estimation error. We find that AVMA maximizes audience value, increases network revenues relative to a variety of benchmarks, and runs fast enough to implement at realistic scales.

The most significant long-run impact of addressing audience externalities would be that networks could charge advertisers for causing audience loss, thereby rewarding them when they select advertising content that reduces viewer switching. Networks would benefit from higher advertiser demand for their product and optimal selection of commercial break length. Advertisers with low-switching ads and high willingness to pay would benefit from a more efficient marketplace with larger commercial audiences available. Viewers would collectively benefit as correcting audience externalities would increase the utility of television.

1.1. Audience Externalities in Television Advertising

Television programs are subsidized by advertisers, but advertisements cause audience loss. Viewer switching, also known as "zapping," is often related to ingrained habits but can also be caused by ad-specific content or stimuli. Any ad that causes a particularly large audience loss reduces the number of viewers left to watch the remainder of the commercial break.

Networks' traditional business model has been to sell ad slots to the highest bidders and assign those ads at random within commercial breaks. In the United States, advertisement sales contracts typically require the network to assign ads to slots within commercial breaks on an "equitable" basis, which is commonly understood to mean quasi-random (Mandese 2004; Downey 2006, §4.1). The first slot in a break ("A position") is the most valuable to advertisers since it typically carries the largest audience. However, with just a few exceptions (e.g., the Super Bowl), this value is seldom priced. Although networks sometimes do price differentially by slot within the break, the price to the advertiser never depends on the potential audience loss its specific commercial can cause. As far as we know, this last observation is true of every country's television advertising marketplace.

This pricing structure limits advertisers' incentives to retain viewers. An advertiser certainly has some incentive to keep people watching: an ad that causes many viewers to switch channels may engender ill will toward that advertiser's brand. But consider a stimulus that could be placed in an ad that simultaneously increases advertising effectiveness by 10% and causes a 5% audience loss (for example, a car dealer screaming at the camera to "break through the clutter"). A rational advertiser will include this stimulus in its ad because the effectiveness multiplier of 1.1 outweighs the audience reduction of 0.95. However, this action causes an audience externality: it reduces the audience remaining to watch commercials in the following ad slots, a negative effect that is not reflected in the advertiser's spot price. Because every advertiser faces similar incentives, even a small externality could compound across multiple slots and lead to an aggregate effect of considerable magnitude.

Thanks to recent technology changes, TV networks now have the potential to resolve these audience externalities. Many companies—including comScore, Kantar, Rentrak, and TiVo—have recently started offering digital set-top box data to estimate or measure program audiences. The largest of these, Rentrak, already measures viewing in more than 20 million television households—nearly 1,000 times more than Nielsen's NTI sample. Early adopters of Rentrak's service mostly include the "long-tail" broadcasters whose audiences are too small to be reported by Nielsen but collectively account for about 15% of TV viewing in the United States (Morgan 2013). These new sources of audience measurements have also been adopted by numerous advertisers. The data have been used to negotiate ad prices, but we are so far unaware of any systematic uses that address audience externalities.[2]

1.2. Relationship to Prior Literature

Audience externalities have not previously been considered in the academic or practitioner literature on television, but they have been studied in the context

---

[2] Digital set-top box data have enabled two other relevant innovations. Several new firms (e.g., Ooyala, Simulmedia, TRA) combine set-top box data on advertising exposures with public records and private databases to improve advertisers' audience purchases. Also, cable companies have developed "addressable advertising" systems to deliver different sets of ads to different segments of households. For example, a single commercial slot could be used to advertise product a to segment 1 and simultaneously advertise product b to segment 2. Though often misunderstood as individualized targeting, these systems actually target groups of households based on geography and demographics (Kumar and Schechner 2009). If households within a target group are heterogeneous in their zapping behavior, the algorithms presented here might be usefully applied at the level of the target group.

of Internet advertising. Google considers the click-through rate of search advertisements to indicate the "quality," or consumer acceptance, of a search advertisement; ads with higher click-through rates get better slots and pay lower prices, all other things equal. Gomes et al. (2009) showed audience externalities to be economically and statistically significant in search advertising. Click-through rates of search ads are conceptually similar to zapping rates for television ads.

A model of audience externalities in search advertising was independently and simultaneously developed by Aggarwal et al. (2008) and Kempe and Mahdian (2008). These authors studied a market in which a search engine sells multiple consumer clicks on a number of consecutive advertisement slots to a group of self-interested advertisers. Advertiser valuations are privately held, but advertisements' click-through rates are known by the search engine. The key result of these papers is that, given a homogeneous audience response to ads, there exists an optimal sorting heuristic for advertisements. This heuristic was combined with dynamic programming to assign advertisements to slots.

This article extends the models of Aggarwal et al. (2008) and Kempe and Mahdian (2008) by replacing their assumption of homogeneous ad response with a set of discrete heterogeneous audience segments. In the presence of heterogeneity, the key insight of Aggarwal et al. (2008) and Kempe and Mahdian (2008)—the global sorting rule—ceases to hold. As a result, it becomes necessary to evaluate many candidate sorting rules in conjunction with a more complicated multisegment dynamic programming algorithm. We also contribute by optimizing commercial break length selection and accommodating slot-specific zapping rates and anticompetition constraints among advertisers.

A second approach to handling audience externalities was developed by Abrams and Schwarz (2008), also in the context of search advertising. Under their model, each advertiser chooses an offering (such as a landing page); different offerings have different trade-offs between nuisance to the consumer and profit to the advertiser. The chosen offering and its nuisance can be observed by the search engine. The key modeling distinction is that the nuisance to the consumer directly translates into a cost to the search engine, in that a higher nuisance reduces consumers' propensity to click on ads in the future, and thus the search engine's future revenue. The main technical contribution of Abrams and Schwarz (2008) is an extension of the Generalized Second-Price (GSP) auction of Edelman et al. (2007) and Varian (2007) to explicitly charge advertisers for nuisance costs and thereby provide incentive for each advertiser to choose the offering that maximizes the joint welfare of the advertiser and search

engine. Stourm and Bax (2013) showed that when ads with higher nuisance costs also provide more value to their advertisers, the GSP extension of Abrams and Schwarz (2008) will improve the advertisers' collective welfare when they are sufficiently numerous. Similar to Aggarwal et al. (2008) and Kempe and Mahdian (2008), these papers assume homogeneity in how audiences respond to ads.

In fact, going back to Rust et al. (1992), the empirical literature on television viewing has been unanimous on the importance of heterogeneity in explaining television audience behavior (Danaher and Mawhinney 2001, Goettler and Shachar 2001, Schweidel and Kent 2010, Shachar and Emerson 2000, Siddarth and Chattopadhyay 1997, Wilbur 2008). Heterogeneity is more important in the context of television signal distribution than search advertising: whereas search ads can, in principle, be targeted to individual users, broadcast and cable delivery systems are designed to serve identical sets of channels to many different consumers simultaneously. There is substantial empirical support for the idea that zapping rates are heterogeneous across viewers and advertising creatives (Teixeira et al. 2010, Wilbur 2013, Wilbur et al. 2008, Woltman Elpers et al. 2003), a result that is replicated in §4 of the present study.[3]

2. Models and Mechanisms

We model consumers' viewing decisions and advertisers' purchasing decisions for one commercial break, defined as J sequential ad slots indexed by j. These slots are to be filled by a subset of n available advertising creatives, which are indexed by a.[4] The ad placed in slot j is denoted by a_j. The network's problem has four aspects:

1. determining J, the number of slots to sell;

2. selecting from the n available ads to fill those slots;

3. ordering the selected ads among the available slots; and

4. pricing each ad to eliminate audience externalities.

The network's choices have to satisfy anticompetition constraints. Television advertising contracts prevent networks from showing direct competitors' ads within the same break, because this leads some consumers to misattribute product a_1's message to

---

[3] A related literature estimated how advertising "clutter" changes advertising effectiveness (e.g., Burke and Srull 1988, Webb and Ray 1979). This article, by contrast, considers how a network might change its business model to optimize the amount and nature of advertising clutter.

[4] Multiple creatives may belong to a single advertiser.

product a_2's ad, and vice versa (Brown and Rothschild 1993, Burke and Srull 1988).[5]

2.1. Viewership Model and Audience Loss

We assume that the audience is divided into S demographic segments $s = 1, \ldots, S$. Segments are most easily defined in terms of demographics that advertisers have traditionally valued ("women 18–49," "men 18–49," etc.).

The audience size at the start of the break is normalized to $\sum_{s=1}^{S} \nu_s = 1$.[6] Here, ν_s represents the proportion of viewers in segment s at the start of the break and is commonly known to the network and its advertisers.

Each viewer in segment s, after being exposed to advertisement a, will continue watching until the next ad slot with a known probability $c_a^{(s)}$. We refer to these parameters as *continuation rates*. If the viewer does not continue watching, he will leave the audience by changing channels or turning off the television, which happens with probability $1 - c_a^{(s)}$. A viewer who continues to the next ad will subsequently make an independent decision whether to continue viewing that ad. If the viewer leaves the audience at any point, we assume that he misses the rest of the break.[7] To employ the models and algorithms proposed here, networks need to know the continuation probabilities. Online Appendix C.7 (available as supplemental material at http://dx.doi.org/10.1287/mksc.2013.0807) discusses how the continuation probabilities could be estimated for new advertisements.

Audience size during the commercial break is influenced not only by ad content but also by exogenous factors such as ingrained ad avoidance habits and audience inflow from other networks. These exogenous factors are slot-specific, as habitual ad avoiders are much more likely to depart during the first slot in the break than later slots. They are also likely to lie outside the network's control in the short run. We model such behavior with slot-specific continuation rates ϕ_j: regardless of the ad content of slot j, the audience at the beginning of slot $j + 1$ will be scaled by a factor ϕ_j.[8] Combining all aspects of the viewing

model, the number of viewers in segment s remaining at the end of slot j is given by

$$P_j^{(s)} = \nu_s \prod_{i=1}^{j} \phi_i c_{a_i}^{(s)}. \tag{1}$$

2.2. Advertisers and Audience Value

Advertisers seek to purchase slots so as to show their ad creatives to the audience. They may value viewers in each segment differently. Each advertisement a produces an advertiser value $\psi_a^{(s)}$ per unit of audience segment s; the parameters $\psi_a^{(s)}$ are private information held by the advertisers. The advertiser thus derives a utility of $\sum_{s=1}^{S} P_j^{(s)} \psi_a^{(s)}$ from having ad a shown in slot j.[9]

Anticompetition constraints are modeled as follows. There exist G disjoint subsets $A_1, \ldots, A_G$ of advertisements for products that compete within markets $g = 1, \ldots, G$. Every advertisement a is contained in exactly one subset A_g; for example, one such subset might include diet cola ads such as those from Diet Coke and Diet Pepsi. Any advertisement a without a direct competitor exists as the lone member of its subset. No feasible solution to the advertisement selection problem may include more than one advertisement from any subset A_g.

Summing over all slots, the total value of the audience to advertisers (equivalently, *advertiser welfare*) from a feasible allocation $a_1, \ldots, a_J$ of ads to the J slots is

$$W(a_1, \ldots, a_J) = \sum_{j=1}^{J} \sum_{s=1}^{S} P_j^{(s)} \cdot \psi_a^{(s)}. \tag{2}$$

Expression (2) explicitly incorporates audience externalities. An advertisement a_j with a low continuation probability $c_{a_j}^{(s)}$ reduces the number of viewers in segment s available to watch the commercials placed in all subsequent slots $j + 1, \ldots, J$, thus decreasing $W(a_1, \ldots, a_J)$.

2.3. Market Design Framework

Because the advertisers' parameters $\psi_a^{(s)}$ are not known to the network, the advertisers communicate their willingness to pay (WTP) by submitting segment-specific bids $b_a^{(s)}$ per unit of audience of each segment s. In principle, these bids may misrepresent the actual willingness to pay, and strategic advertisers would bid nontruthfully if it were in their interest to

---

[5] Anticompetition constraints remove a second type of negative externality among advertisers, competition externalities (e.g., Athey and Ellison 2011, Jehiel et al. 1996).

[6] This normalization comes without loss of generality because all of the subsequent models and algorithms are independent of the units in which audiences are measured.

[7] The analysis focuses on measurable changes in program audience, such as Nielsen's C3 rating. It is understood that a viewer who does not switch channels or power off the television does not necessarily pay attention to the advertisement.

[8] The model allows for $\phi_j > 1$, indicating that more viewers may tune in during a slot than tune away.

[9] Television advertisements typically begin with entertaining content designed to catch the viewer's attention or influence his mood, then end with persuasive selling messages that benefit the marketer (Ferencz and Teixeira 2010). Therefore, we assume that advertisers only value viewers who continue watching until the end of the slot their message is placed in.

do so. Thus, our mechanism is designed so that advertisers will truthfully reveal the values $\psi_a^{(s)}$, as this facilitates efficient selection and ordering of advertisements. The framework for the proposed mechanism is as follows:

1. Advertisers enter bids $b_a^{(s)}$ expressing their willingness to pay per unit of audience from each segment s.[10]

2. The network selects the number of advertising slots, J.[11]

3. The network chooses a feasible selection of ads $a_1, \ldots, a_J$ and shows them to the audience in slots $j = 1, \ldots, J$. After the break airs, the network and advertisers observe segment-specific audience measurements for each slot in the break.

4. Advertisers are then charged. The price for ad a in slot j is calculated as a function of the size of the audience watching that slot and the amount of switching caused by the ad (after controlling for regular slot-specific switching habits).

Four characteristics of this framework differ markedly from the traditional television business model. First, precise audience measurements are required to calculate advertising prices after the ads air so that advertisers are charged according to their actual reach and frequency, as well as the degree to which their ads keep the audience tuned to the channel. These calculations depend on a reliable source of program audience measurements.

Second, the timing differs from current practice. Traditionally, networks require payment for ads prior to airing them. About 80% of national network inventory is purchased in the May "up-front" market for the following September–August television season. Advertisers have long complained about this inflexible timing because it conflicts with standard internal budgeting time frames. The framework described above would reserve payment until shortly after ads air. At that point, audience externalities may be observed and used to calculate ad prices.

Third, the proposed framework sets prices via a Vickrey-Clarke-Groves (VCG) auction (Vickrey 1961, Clarke 1971, Groves 1973) rather than the traditional negotiation-based marketplace. Auctions have recently proven adept at managing multibillion-dollar online advertising markets and can be automated easily. The VCG framework is proposed because it generates truthful bidding in equilibrium and thereby

makes the rest of the mechanism strategy-proof (conditional on the correctness of the model). A primary purpose of the algorithm is to replace the random sorting of ads with an efficient sorting rule. In the absence of truthful bidding, the network would have to make Bayesian inferences about bidders' valuations in order to efficiently sort ads, a process that could become complicated if bidders were to strategically react to the network's process of forming expectations.[12]

Fourth, one might reasonably ask whether the algorithms presented here are overly complex. The complexity comes from the broad range of problems considered. For example, if the problem were restricted to commercial breaks of fixed, short durations, audience retention would be less important, and heuristics might perform well. If the problem were restricted to cases where the network knows advertisers' valuations, the possibility that strategic advertisers could game the network's pricing mechanism would be eliminated. The goal here is to provide an approach that works well in a broad class of problems, but it is this breadth of problems that necessitates the complexity of the solution.

3. The Audience Value Maximization Algorithm

AVMA is the main algorithmic contribution of our article. It searches for an approximately optimal selection and ordering given a feasible set of potential advertisements, then determines VCG payments given that selection and ordering. To preserve readability, we first present the algorithm conditional on a fixed number of slots and no slot-specific continuation rates or anticompetition constraints. We then subsequently show how to accommodate these factors.

In the proposed framework, the television network seeks to maximize the value it creates for its advertisers in order to maximize its own ability to extract revenues from them. Advertising prices are then determined using the VCG mechanism (Clarke 1971, Groves 1973, Vickrey 1961). VCG auctions give bidders strictly dominant strategies to bid truthfully;

---

[10] It seems reasonable to expect delays of up to a few days between entering bids and determining slot order, as well as between observing audience externalities and calculating payments. This stands in contrast to Internet advertising auctions, which run nearly continuously throughout the day.

[11] The choice of J may be restricted by network policy, program constraints, or government regulations.

[12] An important, related question is whether the VCG auction maximizes expected seller revenues. This question remains open. In contrast to the classic type of analysis by Myerson (1981), the heterogeneous audience externalities in this setting eliminate the anonymity that characterizes most auction analysis. AVMA could be viewed as a means of finding a set of consistent, efficient, feasible allocations; systematically evaluating that set; and selecting the revenue-maximizing allocation conditional on a particular payment mechanism. From this standpoint, the desirability of the strategy-proof property becomes clear, as the seller would have trouble characterizing which allocations are consistent and efficient without truthful bids.

hence, we assume truthful bidding for the remainder of this section, as it is consistent with advertisers' rational self-interest.

The most difficult part of the algorithm is finding a feasible advertisement selection and ordering. These two decisions are heavily intertwined in the following sense: the optimal relative ordering of any two ads a and a' depends not only on parameters of these two ads (their continuation rates and willingness to pay) but also on which other ads are shown. Conversely, the correct set of ads to show depends on the order in which they will be shown.[13] This difficulty is introduced by the inclusion of audience heterogeneity in the model; in the homogeneous-audience version of the model, the optimal ordering of ads does not depend on which ads are selected.

3.1. Advertisement Selection, Ordering, and Pricing

The main intuition behind the selection and ordering algorithms is presented here; the full specification is given in Appendix A. The basic idea is to consider many different orderings for the ads and then use dynamic programming to calculate the optimal selection of ads for each ordering. Each ordering considers not just a single segment but a combination of segments. Finally, from among all these different orderings, the one maximizing audience value is selected.

The main idea behind AVMA is most easily illustrated geometrically. Each advertisement a is associated with a point $\mathbf{q}_a \in \mathbb{R}_+^S$, with coordinates $q_a^{(s)} = (b_a^{(s)} c_a^{(s)})/(1 - c_a^{(s)})$. Thus, ads will be farther from the origin in a dimension s (which represents segment s) when they retain more viewers in that segment and value viewers in that segment more highly, as illustrated in Figure 1 for $S = 2$. The selection rule then generally favors the advertisements farther from the origin for its best slots; these advertisements correspond to the network's best customers (in terms of willingness to pay and ability to retain audience).

[13] To see this, suppose that we have $S = 2$ segments with $\nu_1 = \nu_2 = 0.5$, three slots, and four ads. Two ads have continuation rate vectors $\mathbf{c}_1 = \mathbf{c}_2 = (1, 0)$, and the other two ads have $\mathbf{c}_3 = \mathbf{c}_4 = (0, 1)$. In other words, the first two ads retain all viewers from segment 1 and lose all viewers from segment 2, whereas the other two ads retain all viewers from segment 2 and lose all viewers from segment 1. These figures are extreme, but they only serve to illustrate a simple point. The willingness to pay for all ads and segments is 1. If ads 1, 2, and 3 are selected, then ad 3 must always be placed last; if either ad 1 or ad 2 were to appear after ad 3, then only the first slot would produce any value for any advertiser. On the other hand, if ads 1, 3, and 4 are selected, then ad 1 must be shown last. Once ad 1 has been shown, no viewers remain in segment 2; therefore, ad 3 or ad 4 will be worthless any time it follows ad 1. As a result, the ideal ordering between ads 1 and 3 depends on whether the third ad selected is ad 2 or ad 4.

Figure 1   Two-Segment Illustration
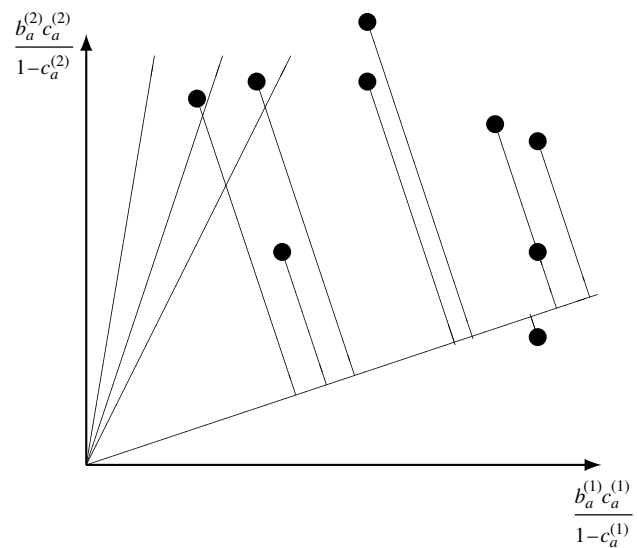
More formally, AVMA considers a set of lines spanning the space. Each line represents a candidate sorting rule. The algorithm then works as follows:

Step 1. For a line in the positive orthant, project all advertisements onto that line and run a multisegment dynamic programming algorithm to calculate the optimal advertisement selection and maximal audience value available along that line.

Step 2. Repeat Step (1) for many lines in the positive orthant, evenly spread over the space of different weightings of segment importance.

Step 3. Select the line and corresponding ad selection that produces the maximum audience value.

For any given ordering of ads, the dynamic program chooses the optimal trade-off between monetization and audience retention for each slot. As the break proceeds, this trade-off changes: retention becomes less important relative to monetization. The dynamic program figures out the best way to make these trade-offs given the specific set of advertisements available.

The algorithm uses approximations in two places. First, the parameterization of lines in Step (2) will roughly cover all available directions, and thus weightings of segments; however, because there are infinitely many lines that could be evaluated, the set of lines explored is necessarily discretized, for example, in increments of one degree in each dimension. Second, the audience remaining in each segment after each slot is discretized to enable the dynamic program to quickly compute the welfare-maximizing selection of ads for a given ordering rule.

Advertisement prices are computed using standard VCG payments, charging a price of $\sum_s b_a^{(s)} P_j^{(s)} - (W - W_{-a})$ for the advertising creative a when it is

placed in slot j. W_{-a} denotes the maximum audience value available when creative a is not aired. By construction, this payment never exceeds a's aggregate willingness to pay, $\sum_s b_a^{(s)} P_j^{(s)}$. Furthermore, by the standard VCG pricing argument, so long as the solution W as well as all the W_{-a} are computed optimally, each advertiser has a dominant strategy to reveal its true willingness to pay.

3.2. Choosing the Break Length

The network can maximize its advertising revenue by running Algorithms 1 and 2 in Appendix A for every feasible choice of break length J, then selecting the number of slots that maximizes the advertising payments it receives. Total audience value W can only increase with J because more advertisements are served when the network increases the number of slots available. However, past some point, the network's revenue must decrease with J. This is because each advertisement's VCG payment is a decreasing function of $W - W_{-a}$, the change in aggregate audience value when that advertisement is replaced with the best available excluded advertisement. With each increase in J, fewer and less desirable ads are available for inclusion. Since replacing any included ad a with an excluded advertisement then causes a larger decrease in audience value, the payment for each included ad must decrease.

3.3. Slot-Specific Continuation Rates

Figure 2 presents proportional audience inflow and outflow, averaged across all eight-slot breaks observed in the viewing data described in the next section. Outflow spikes in the first slot, remains high in the second, and then lies flat across all remaining slots in the break. Inflow and outflow both average 2.9% for slots 3–8, approximately offsetting each other. Note that no spike in inflow is observed in the final slot of the break. The data show highly similar patterns for breaks of all common lengths J.

Figure 2   Proportional Tune-in and Tune-away by Slot
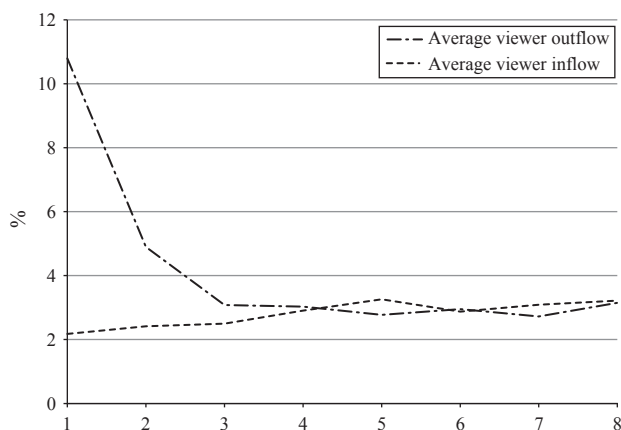
These data imply that slot-specific continuation probabilities ϕ_j lie strictly below 1 for the first two slots in the break but are approximately equal to 1 in all remaining slots. Let N be the number of initial slots for which $\phi_j \neq 1$. For small N, it is practically feasible to try all N-tuples of ads in the first N slots. For each possible assignment and order of N ads to these N slots, the algorithms in Appendix A can be run for the remaining $J - N$ slots and $n - N$ ads, conditioned on the audience loss caused by the first N slots. Having tried all these selections, the best one can then be retained.

3.4. Anticompetition Constraints

AVMA takes as an input a feasible set of potential advertisements, i.e., one that does not violate any anticompetition constraints. There are three approaches to find such a feasible set, with the optimal solution depending on the sizes of the anticompetition sets of advertisements A_g.

The simplest way to approach the issue is to simply run AVMA for all possible sets of advertisements and only retain a solution (in line 13 of Algorithm 1 in Appendix A) when it satisfies the anticompetition constraints. This approach is simple and likely to work in settings where n is small or the number of anticompetition sets G is large relative to n.

A second approach would be to treat the solution to the anticompetition constraint problem as an outer layer and search over the space of feasible sets of advertisements using a technique such as simulated annealing (Kirkpatrick et al. 1983). The outer loop selects as its solution a set of $n_0 \leq n$ potential advertisements that are conflict-free in the chosen model. This subset of ads would still contain more ads than there are slots in the break. AVMA can then be used as the critical routine to evaluate the maximal audience value that can be obtained from the subset of candidate advertisements, which helps the search loop decide which solution to retain or improve. This approach decouples the two difficulties of avoiding conflicts and selecting and ordering ads. AVMA is a good algorithm for the latter, and standard techniques can be applied for the former.

The third approach, which should be used in combination with either of the first two, would be to simplify the anticompetition constraints by eliminating dominated advertisements. For any two competing ads a and a', if $b_a^{(s)} \geq b_{a'}^{(s)}$ and $c_a^{(s)} \geq c_{a'}^{(s)}$ for all s, then ad a' can be eliminated, because it will never be selected when ad a is available as well.

4. Estimation

In this section, we estimate the primitives of the model: heterogeneous continuation probabilities, segment sizes, and advertisers' willingness to pay. Our

purpose is twofold: one is to learn about advertising avoidance behavior, and the other is to generate an environment for realistic assessment of AVMA performance and feasibility relative to alternative solutions.
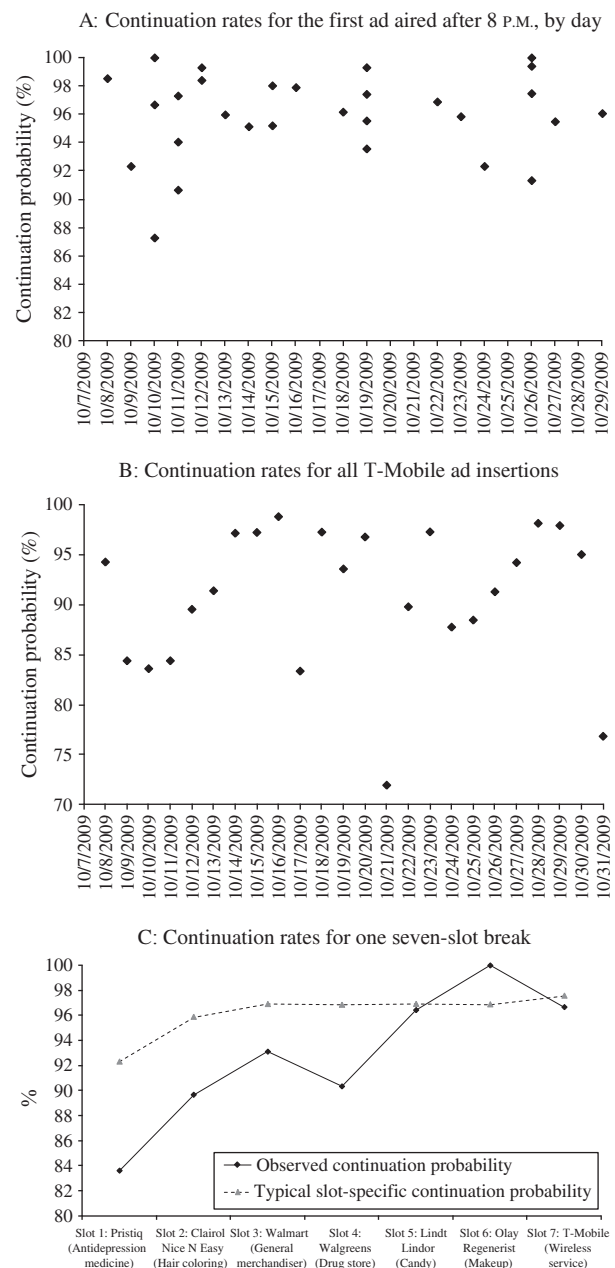
4.1. Data Sets

The analysis of viewer behavior combines two data sets. The viewing data were provided by executives at Kantar for the explicit purpose of testing AVMA performance. The data record all non-recorded television-viewing sessions in October 2009 for 266,000 set-top boxes (STBs) connected to a digital cable television system. For each set-top box in each second of the month, the data record whether the box was in use and, if so, the channel to which it was tuned. The data are completely anonymous; U.S. law prohibits cable operators from providing personally identifying information to any third party without subscribers' explicit consent.

Viewing data are supplemented with advertising data collected by Kantar's Stradegy database, the leading source of competitive advertising intelligence. For each advertisement aired by a national cable network, the data contain the ad start time, its duration, brand and campaign identifiers, and its estimated price. The data show no systematic pattern of placing particular ads in particular slots.

The empirical analysis focuses on a single cable network, Turner Network Television (TNT). The single-network focus is motivated by the observations that this is the level at which a network business model decision would be made and that media conglomerates do not typically incorporate advertiser substitution across networks when setting advertising prices. TNT is one of the highest-revenue networks in the data and typically airs recorded programs, which facilitates AVMA implementation by making commercial breaks predictable and controllable. The model is estimated using household-level data to maximize asymptotic efficiency. If all set-top boxes and all ads were used, the parameter space would grow unreasonably large. Accordingly, the analysis focuses on prime time (8–11 P.M. Monday–Saturday and 7–11 P.M. Sunday), the 25 most aired advertisements, and a set of viewers that collectively accounted for 90% of network viewing.

Prior literature identified three sets of factors that may influence observed rates of advertising avoidance: ad content, environmental factors, and audience heterogeneity. Ad content refers to the specific audio and visual stimuli recorded within a commercial. Environmental factors include break or slot effects as well as the type of program during which the break occurs. Audience factors include differences in habits as well as how long the individual has been watching the channel. Figure 3 substantiates that continuation rates depend on all three sets of factors.

Figure 3   Model-Free Evidence of Switching Heterogeneity

A: Continuation rates for the first ad aired after 8 P.M., by day

B: Continuation rates for all T-Mobile ad insertions

C: Continuation rates for one seven-slot break

Panel A of Figure 3 shows the continuation rates observed for the *first* ad aired on TNT after 8 P.M. on each day of the estimation sample. By examining the first ad after 8 P.M., we hold break and slot environmental factors constant while audience factors and ad creatives change across days. The data show substantial variation, with continuation rates ranging from 72% to 99%. The average first ad after 8 P.M. retained 90.9% of its audience with a standard deviation of seven percentage points.

Panel B of Figure 3 shows the observed continuation rates for all prime-time airings of the most frequently occurring ad creative in the sample (T-Mobile

wireless service). Continuation rates range from 87% to 100% with no discernible trend over time. The ad aired three times on October 10, 2011, with substantial variation in audience retention even within a single evening; its continuation rates were 87%, 97%, and 100%, respectively.

Panel C of Figure 3 displays the continuation rates observed within a particular commercial break in the sample, and the typical continuation rates observed across all breaks in the sample. By focusing on a particular break, we are holding initial audience factors constant, although slot and ad creative factors change over slots. This break was chosen because it contained 3 of the 25 most aired ad creatives. The first slot in the break was used to advertise an antidepression medication and retained just 84% of its audience. This was an unusually high dropout rate, as the "A" position normally retains 93% of its audience. Continuation rates are nonmonotonic over the course of the break, rising to 90% and 93% in the following two slots, then falling to 90% in the fourth slot before rising again in the next two slots and finally falling again in the last slot of the break. Factoring in tune-in and tune-away for each slot of this break, the network's overall audience was 67% smaller at the end of the break than at the start.

In summary, Figure 3 shows that ad creative, environmental factors, and audience factors jointly influence continuation rates. The goal of the empirical model is to estimate the impact of advertising creatives on continuation rates, separate from the influence of environmental and audience factors.

A question might naturally be raised about how ad-specific effects can be identified separately from slot-specific tune-away rates. Each slot-specific tune-away rate is identified by that slot's mean tune-away rate. Each ad's continuation rate parameter is identified by deviations from typical slot-specific tune-away rates corresponding to placement of the ad in question. There is no apparent correspondence between advertising creatives and ad placements in the data.

The data are well suited to estimate continuation rates, but they do have some limitations, most of which are shared by some or all published studies in this area. To remove possibly latent viewing sessions, the data provider drops all sessions in which a viewer went more than two hours without using the remote control, suggesting a possible downward bias in continuation rate estimates. Digital video recorder (DVR) usage is unobserved, but it is known that even households with DVRs spend large amounts of time watching non-recorded television (Bronnenberg et al. 2010, Zigmond et al. 2009). Furthermore, the data come from a particular digital cable television system rather than a nationally representative panel of television viewers. Only national cable network ads are

observed; to the best of our knowledge, no syndicated data provider records local cable ads or advertisements for upcoming cable network programs ("promos" or "tune-ins"). These imperfections in the data may affect the estimation results, as they did in other studies that used similar data sets. However, the same environment will be used to evaluate all candidate algorithms, so performance comparisons might not be overly biased by the data imperfections. We see no reason to believe that the possible biases would favor one algorithm over another.

4.2. Empirical Findings About Viewer Behavior

We estimate a discrete choice model with heterogeneous segments to isolate the effects of advertising creative on continuation rates. The STB data are anonymized, rendering demographics unavailable, so set-top boxes' observed behaviors ("usographics") were used to predict the STBs' segment memberships, as in Chintagunta and Gupta (1994). Two usographics were found to predict segment membership: television viewing propensity and ad tune-away propensity.[14]

The model estimates each segment's tendency to continue viewing each of the 25 included advertising creatives, controlling for viewing session factors, past exposures, program genre, break and slot factors, and weekday and half-hour within prime time. Appendix B presents the model of viewing behavior in full detail.

Two-, three-, and four-segment models were estimated. Table 1 shows that the Bayesian information criterion (BIC) favors the two-segment model over both simpler and more complex alternatives, so for the remainder of this subsection, we discuss estimates for the two-segment model only. All of the statistically significant results in the two-segment model were replicated by the three- and four-segment models' parameter estimates, and the richer models did not produce any qualitatively different behavioral findings.

Table 2 shows that the two viewing segments differ markedly in their observed switching behaviors. The two segments can be roughly characterized as frequent switchers (Segment 1) and couch potatoes (Segment 2).

Set-top boxes that watch *more* television and avoid *fewer* ads are more likely to be assigned to Segment 2

---

[14] To avoid simultaneity, no data related to the viewing of the focal network (TNT) were used in constructing these variables. We also estimated models including the proportion of viewing time spent on each major competing network (e.g., ESPN, Oxygen) and the proportion of viewing done on each weekday. Estimation results did not suggest that these additional features were useful in assigning households to segments.

Table 1   Segmentation Results

			Number of viewer segments					
	$S = 1$		$S = 2$		$S = 3$		$S = 4$	
Log likelihood	−3,352		−3,038		−2,961		−2,823	
No. of parameters	50		102		154		206	
BIC	7,218		7,124		7,504		7,761	
Effects on segment membership probability	Est.	T-stat.	Est.	T-stat.	Est.	T-stat.	Est.	T-stat.
Viewing propensity								
Effect on Segment 1 membership			−0.005	8.8**	−0.003	2.4*	0.000	0.0
Effect on Segment 2 membership					−0.005	6.2**	0.003	2.1*
Effect on Segment 3 membership							−0.003	2.2*
Tune-away propensity								
Effect on Segment 1 membership			0.210	7.1**	0.136	2.5*	0.000	0.0
Effect on Segment 2 membership					0.311	6.9**	−0.156	1.8
Effect on Segment 3 membership							0.270	4.8**
Estimated segment size								
Segment 2				47%		27%		19%
Segment 3						41%		25%
Segment 4								37%

Note. Number of observations = 28,867.
  *Statistically significant at the 95% confidence level; **statistically significant at the 99% confidence level.

Table 2   Viewer Model Parameter Estimates

	Two-segment viewer model			
	Segment 1		Segment 2	
Predictor	Est.	T-stat.	Est.	T-stat.
Tuned in during slot	−0.7	4.9**	1.7	5.1**
Consecutive minutes watching	0.0	17.5**	2.4	7.1**
Previous exposure of h to a in past week	0.0	0.5	0.1	0.6
Genre: Feature film	−0.4	2.2*	−1.0	1.6
Genre: Police/suspense/mystery	−0.1	0.7	−0.9	2.1*
Genre: Pro basketball game	1.2	3.6**	−0.6	0.8
First break in program	0.0	0.1	0.4	0.9
Last break in program	0.0	0.2	−0.2	0.5
Second slot of break	1.1	6.1**	−0.2	0.5
Third slot of break	1.3	6.5**	0.1	0.1
Fourth slot of break	1.6	8.7**	−0.6	1.0
Fifth slot of break	1.6	7.1**	−1.1	1.6
Sixth slot of break	1.3	5.8**	−0.5	0.8
Seventh+ slot of break	−0.4	2.2*	0.1	0.2
Tuesday	−0.1	0.7	0.2	0.2
Wednesday	0.3	1.3	0.7	0.9
Thursday	0.3	1.1	0.1	0.1
Friday	0.4	1.9	0.2	0.3
Saturday	0.2	1.1	0.3	0.5
Sunday	0.3	1.1	0.5	0.7
8:30 P.M.–9 P.M.	−0.3	1.8	0.6	1.2
9 P.M.–9:30 P.M.	−0.1	0.7	0.3	0.6
9:30 P.M.–10 P.M.	−0.3	1.4	1.5	2.6*
10 P.M.–10:30 P.M.	−0.3	1.1	0.0	0.1
10:30 P.M.–11 P.M.	−0.2	0.8	0.9	1.4

  *Statistically significant at the 95% confidence level; **statistically significant at the 99% confidence level.

("couch potatoes"). Couch potatoes show two interesting behaviors related to ad avoidance. First, they are no more likely to zap an ad in the first slot than in later slots in the break. Second, if they tune in while an ad is in progress, they tend to not zap before the end of the ad slot.[15] Viewers who watch *less* television and zap *more* ads are more likely to be characterized as "frequent switchers." These viewers are far more likely to zap during the first slot of the break than during later slots. Unlike couch potatoes, if they tune to the channel while an ad is in progress, they are more likely to switch away again before the ad finishes.

The two segments also differ in how program genre influences switching. Frequent switchers are *more* likely to zap an ad when it appears during a movie and *less* likely to zap ads during basketball games. The police/suspense/mystery genre had no apparent effect on zapping for this segment. On the other hand, couch potatoes were *more* likely to zap during a police/suspense/mystery program but show no apparent effects of movie and basketball genres on their zapping behavior.[16] These results confirm the importance of environmental factors in predicting ad avoidance.

The segments exhibited fewer differences in other factors that might have influenced zapping probabilities. The amount of time spent watching the channel prior to the current ad break is a strong predictor

[15] Of course, not zapping does not necessarily imply that the viewer watched the ad.

[16] These genres are nonexhaustive, but they reflect the range of programs offered in prime time by the focal network (TNT).

Table 3   Continuation Rate and Willingness-to-Pay Estimates

a	Advertising brand // Campaign name	No. of airings on TNT	Continuation rates				Advertiser WTP ($)	
			Segment 1		Segment 2			
			Estimate	SE	Estimate	SE	Estimate	SE
1	T-Mobile // *Saturday Night Live* Cast Shares Phone	29	0.968	0.002	0.954	0.002	29.78	3.88
2	Jameson Irish Whiskey // John Jameson	16	0.948	0.002	0.919	0.002	30.06	6.84
3	Chase Sapphire Card // M. Culver Uses Points to Buy Dress	13	0.952	0.001	0.932	0.001	19.49	2.18
4	Walmart // Make Your Home Picture Perfect This Fall	13	0.938	0.003	0.898	0.003	29.25	4.63
5	Quiznos Double Cheese Cheesesteak // Men Sit in Hillbilly Hot Tub	13	0.958	0.001	0.927	0.002	19.26	1.86
6	JCPenney // Affordable, Unforgettable Fall	12	0.953	0.002	0.932	0.002	22.75	4.75
7	*Cirque du Freak: The Vampire's Assistant* // Have You Ever Dreamed of . . .	12	0.952	0.002	0.923	0.002	20.28	3.11
8	Lindt Lindor // Chefs Pouring Chocolate	11	0.947	0.003	0.908	0.003	19.98	2.28
9	Kenmore Elite He5T // How Much Can You Get?	11	0.936	0.004	0.889	0.004	25.48	4.25
10	Red Lobster Restaurant // Endless Shrimp: Wood-Grilled Teriyaki	10	0.943	0.002	0.911	0.002	19.44	3.15
11	PetSmart Center // Favorite Things Sale 26	10	0.926	0.004	0.875	0.004	15.27	1.85
12	Jim Beam // Men Use Puppies to Get Girls	10	0.933	0.004	0.898	0.004	32.34	6.30
13	Acura TL Leasing // Precision Is Power	10	0.947	0.001	0.919	0.002	33.79	8.75
14	Dairy Queen Sweet Deals // The Lips Note Deals	10	0.948	0.002	0.906	0.002	9.07	2.16
15	Battlestar Galactica Series // Exclusively On DVD & Blu-Ray . . .	9	0.927	0.006	0.890	0.006	39.47	5.86
16	Ashley Furniture Home Store // Home for The Holidays Event 7	9	0.939	0.004	0.898	0.004	17.50	2.22
17	Hyundai // Information More Important Than Money	9	0.941	0.004	0.921	0.004	17.54	2.67
18	Arby's $5.01 Combos // Introducing the New $5.01 Combos	9	0.943	0.001	0.916	0.001	11.62	2.73
19	Autotrader.com // No Title Assigned	9	0.946	0.002	0.919	0.003	33.70	6.42
20	*Men Who Stare at Goats* // On 11/6, Clooney, Bridges, McGregor . . .	9	0.937	0.008	0.911	0.021	35.45	5.31
21	Pristiq // Woman Is Depressed	9	0.918	0.004	0.856	0.004	24.96	5.38
22	Target // Audrey, Kim, and Karen: Frugalista to Shop	8	0.933	0.003	0.887	0.004	25.19	4.03
23	Walmart // Dining Out Can Cost You 50	8	0.924	0.003	0.897	0.003	24.63	4.19
24	Michelin // Evil Gas Pump Captures a Town	8	0.892	0.007	0.882	0.007	19.15	2.44
25	Acura TSX Leasing // It's Not Always Easy/299	8	0.931	0.002	0.906	0.002	24.00	2.36

of continued ad viewing for both segments because it likely correlates with viewer involvement with the current programming (or with the viewer not paying attention to the TV), but its effect is much stronger for couch potatoes. Past exposures to the advertising creative had no significant effect on switching. Ad avoidance tendency is no higher or lower during the first or last breaks of a program than during the intervening breaks. There are few discernible effects of weekdays or half-hours within the break.

Table 3 displays the estimated continuation rates for each ad in each segment. Standard errors are calculated by bootstrapping from the asymptotic joint distribution of the parameter set. There is substantial heterogeneity in switching probabilities across both segments and ads. The switching behavior of viewers in Segment 1 is more regularly attributable to *slot-specific* factors than that of viewers in Segment 2. On the other hand, *ad-creative factors* play an important role in predicting both segments' zapping behavior. The model estimates that about 6% of switching done by Segment 1 and 9.7% of switching done by Segment 2 is attributable to advertising creatives.

The ad with the highest estimated continuation rate in both segments was a T-Mobile ad that featured former *Saturday Night Live* cast members. In this ad, Chevy Chase, Molly Shannon, Dana Carvey, and Darrell Hammond take turns providing comedic inter-

ludes and handing the phone to each other. The ad with the lowest estimated continuation rate was for a depression medicine called Pristiq. Kantar assigned this ad a title of "Woman Is Depressed." In it, a woman compares herself to a wind-up doll that has run out of energy. We conclude that the estimated continuation rates have a reasonable degree of face validity.

4.3. Empirical Findings About Advertiser Willingness to Pay

Advertisers make trade-offs when choosing whether to purchase ad slots on TNT or on other networks. Therefore, advertisers' observed WTP on *other* networks should indicate their willingness to pay for audiences on TNT.

We were able to obtain national audience estimates for programs that aired on the biggest four broadcast networks (ABC, CBS, FOX, and NBC) between October and December 2009 from Nielsen Media Research. We combined these with Kantar ad price data[17] to calculate advertisers' cost per thousand (CPM) households for insertions on broadcast networks. Under the assumption that an advertiser's WTP per unit of audience is equal across networks, the estimated WTP for

---

[17] The limitations of the ad price data are well known: they are estimated at the program level and do not vary across advertisers or breaks within a program.

each of the 25 advertisements is the average CPM over its ad spots.[18]

Table 3 shows the advertisers' WTP estimates. They range from \$9.07 to \$39.50 per thousand households, with an average of \$24 per thousand households and a standard deviation of about \$4. These estimates will be used in the comparison of the algorithms.

Because the STB viewing data are not representative of the national viewing population, we are not able to estimate how advertisers' willingness to pay varies across audience segments. Consequently, the algorithm-testing environment will assume that advertiser WTP does not vary by segment. This assumption is conservative, because it will reduce AVMA performance relative to simpler alternatives that are not designed to accommodate that heterogeneity.

Overall, there are five types of heterogeneity that matter for network revenues: (1) heterogeneity in tune-away across ad creatives, (2) heterogeneity in tune-away across viewing segments, (3) heterogeneity in advertisers' willingness to pay for viewers, (4) heterogeneity in segments' propensities to tune away from specific ads, and (5) heterogeneity in advertisers' willingness to pay for particular viewer segments. The data have revealed substantial amounts of types (1)–(3) and a limited amount of type (4); limitations in the data prevent us from estimating type (5). The next section investigates how much the network might gain from exploiting these various sources of heterogeneity using a smarter advertising sales algorithm.

5. AVMA Performance and Feasibility

The logic behind AVMA is to create audience value for advertisers and then to extract value for the network. In this section, we present a series of simulations to evaluate the extent to which AVMA accomplishes these goals. All simulations are based on estimates of continuation rates and advertisers' WTP parameters; they explicitly incorporate estimation error.

Section 5.1 shows, for various small combinations of n and J, that AVMA typically maximizes audience value. Section 5.2 endogenizes break length and finds that AVMA produces greater revenues for the network than five alternatives. On average, it increases audience value by 12% and network revenues by 10% over traditional advertisement selection and sorting. Section 5.3 proves that AVMA running time is quick enough to easily scale to the problem

sizes facing the typical network. Section 5.4 closes with a discussion of when AVMA would likely *not* be the best algorithm available.

5.1. Absolute Performance Assessment of AVMA

This section quantifies the degree to which AVMA's advertisement sorting and selection maximizes audience value to advertisers. It does so by comparing the AVMA audience value to the globally optimal audience value. This optimum was computed through brute force: we considered every possible subset of J out of n ads and all available orderings for any such subset, retaining the best audience value available from any selection/ordering possibility.

To incorporate estimation error, we drew 100 times from the asymptotic joint distributions of the parameter estimates. For each draw, we ran both AVMA and the brute-force algorithm. We evaluate the algorithms by comparing the averages and standard deviations of their performance across draws.

The cost of finding the global optimum, given S segments, J slots, and n potential advertisements, is proportional to SJn^J; it therefore increases very rapidly in the size of the problem considered.[19] As a result, this comparison is only feasible for relatively small problems.

Table 4 compares audience value under AVMA to the global optimum for a variety of scenarios. Two conclusions emerge. First, AVMA produces audience value that matches the optimal audience value almost perfectly. Second, it appears that AVMA's ability to find the optimal selection and ordering of ads is not very sensitive to specific values of S, J, and n.

In a few cells in Table 4, AMVA performance is slightly worse than the optimum. As explained in §3.1, AVMA relies on two approximation parameters that can be adjusted to trade off computation time with performance. The results in Table 4 are based on a dynamic programming accuracy parameter of $\delta = 0.07$ and a number of lines $K = 15$. In fact, for sufficiently large K and sufficiently small δ, AVMA has found the global optimum in all settings we have investigated. We quantify this performance/cost trade-off in §5.3.

5.2. Assessing AVMA Relative to Alternatives

AVMA does well at creating value for advertisers, but the question remains: Will networks have an incentive to adopt it? We identify five benchmark algorithms and compare their performance to AVMA for two

---

[18] Advertiser WTP estimation was motivated by our desire for realistic simulation inputs, not inference. The average number of insertions on broadcast networks is about 30 per advertiser. Standard errors were calculated by bootstrapping out of the ad insertions.

[19] For example, in our experiments, the time required to find the optimal solution for a single draw of the parameter estimates rose from less than a minute in the $\{S = 2, J = 8, n = 11\}$ scenario to 21 hours for the $\{S = 2, J = 8, n = 20\}$ scenario. This cost must then be multiplied by the number of draws taken from the asymptotic distribution of the parameter estimates.

Table 4   AVMA Performance on Audience Value

	AVMA Avg. (SD)	Optimum Avg. (SD)		AVMA Avg. (SD)	Optimum Avg. (SD)
Two-segment model, holding slots fixed at $J = 8$			Two-segment model, holding ads fixed at $n = 13$		
$n = 11$	208 (15)	208 (15)	$J = 5$	150 (11)	150 (11)
$n = 12$	206 (12)	207 (13)	$J = 6$	172 (9)	172 (9)
$n = 13$	204 (12)	204 (12)	$J = 7$	188 (10)	188 (10)
$n = 14$	206 (10)	206 (10)	$J = 8$	204 (12)	204 (12)
$n = 15$	204 (14)	204 (14)			
$n = 16$	204 (14)	204 (14)			
Three-segment model, holding slots fixed at $J = 6$			Three-segment model, holding ads fixed at $n = 13$		
$n = 11$	182 (9)	183 (9)	$J = 5$	164 (16)	165 (15)
$n = 12$	185 (11)	186 (11)	$J = 6$	183 (12)	184 (12)
$n = 13$	183 (12)	184 (12)			

Note. Standard deviations are shown in parentheses.

purposes. One is to quantify the gains in network revenue available from correcting audience externalities. The other is to offer insight into how the components of AVMA interact to produce those gains. Each benchmark algorithm is a unique combination of goals and the number of segments considered.

5.2.1. Benchmark Algorithm Goals. The network has two objectives in each slot j in a commercial break.

Slot monetization: Maximize the immediate revenue derived from slot j alone by selling the slot to the highest bidder.

Audience retention: Maximize audience retention in slot j by selling the slot to the advertisement with the highest continuation probabilities. This increases the number of "eyeballs" to be sold to advertisers in subsequent slots.

The parameter estimates in the previous section show that these two goals are often in conflict. Some advertisements with a high willingness to pay have low continuation probabilities, whereas some advertisements with high continuation probabilities have a low willingness to pay.

As discussed in §1.1, television networks have traditionally sold advertisements to the highest bidders and ordered them randomly within the break. This suggests a focus on the slot monetization goal with no consideration given to the audience retention goal.

A different approach would assign some constant trade-off between the two goals. Then, a sorting heuristic could rank ads as a function of both their continuation probabilities and their bids and sell slots to the first J advertisements in the ranking.

However, to maximize revenue, the network must allow this trade-off between goals to change over the course of the commercial break. The importance of the audience retention goal decreases over time, but the importance of slot monetization stays constant. Thus, the optimal balance between the two goals tips toward the slot monetization goal as the break proceeds. AVMA uses dynamic programming to find the best possible trade-off between these two goals in every slot of the break.

The first component used to specify algorithm alternatives relates to their goals.

Traditional sorting and selection: Select ads according to their bids and order them randomly.[20]

Sorting only: Rank ads by a single criterion; then select the highest-ranking advertisements according to that criterion. The criterion used below is the one found to be optimal for a single segment by Aggarwal et al. (2008) and Kempe and Mahdian (2008)—namely, $(b_a^{(s)} c_a^{(s)})/(1 - c_a^{(s)})$. Note that this rule does not maximize audience retention; otherwise, it would simply order advertisements by $c_a^{(s)}$. Rather, it offers a constant trade-off between the slot monetization goal and the audience retention goal without allowing that trade-off to change over the course of the break.

---

[20] In each simulation of the traditional practice, we ran the algorithm 1,000 times; we report the average result of these runs.

Table 5   **Summary of Benchmark Algorithms**

Algorithm	Advertisement sorting	Advertisement selection
One segment		
Traditional sorting and selection	Random	Top J ads by bids
Homogeneous sorting only	Order by $b_a c_a / (1 - c_a)$	Top J ads by sorting criterion
Homogeneous sorting and DP	Order by $b_a c_a / (1 - c_a)$	Single-segment dynamic program
Two segments		
Traditional sorting and selection	Random	According to highest bids
AVMA sorting only	Many candidate sorting rules considered; best one is retained (see §3.1)	Top J ads by best available sorting criterion
AVMA	Many candidate sorting rules considered; best one is retained (see §3.1)	Multisegment dynamic program run for each candidate sorting rule

Sorting and dynamic programming: Use dynamic programming (DP) to optimally balance slot monetization and audience retention in every slot in the break (as in AVMA).

5.2.2. Benchmark Algorithm Heterogeneity. The second component varied across benchmark algorithms is the number of segments they consider. The importance of heterogeneity in television viewing was demonstrated by the result that a two-segment model offered the best fit to the viewing data and by numerous prior studies reviewed in §1.2. We compare AVMA's performance to algorithms designed for a homogeneous audience to quantify the effect of accounting for audience heterogeneity on algorithm performance.

We use a common empirical model to simulate the choices of both one- and two-segment algorithms. This eliminates the empirical model as a confounding factor so any differences in algorithm performance can be attributed solely to the algorithms themselves. Simulations of two-segment algorithm choices are based on draws from the asymptotic distributions of segment sizes (ν_s), segment-specific continuation rates ($c_a^{(s)}$), and advertiser WTP (ψ_a) parameters reported in Tables 1–3. For single-segment algorithms, advertisement selection and ordering decisions are determined by using the following aggregate continuation probabilities,

$$c_a = \sum_{s=1}^{S} \nu_s c_a^{(s)}, \qquad (3)$$

in place of segment-specific parameters $c_a^{(s)}$.[21] Therefore, the continuation rates used by the one-segment and two-segment algorithms are equal in their aggregate market response, but the inputs to the two-segment algorithms give full consideration to viewer heterogeneity, whereas the one-segment algorithms

do not consider segments' differential responses when selecting and sorting advertisements.

5.2.3. Summary of Benchmark Algorithms and Predictions. Six algorithms were compared: AVMA and five others. Table 5 summarizes these benchmarks and the differences between the advertisement sorting and selection rules they employ.

To give maximal weight to network incentives, we incorporate optimal commercial break length selection into all algorithms. For each draw of continuation probabilities and advertiser WTP parameters, and for each algorithm considered, the algorithm computes the selection and ordering of ads for every break length between $J = 4$ and $J = 12$ and then retains the break length maximizing network revenues.

For the single-segment algorithms, advertisement selection and ordering is performed using the c_a parameters defined in Equation (3). The network revenue from those selection and ordering decisions is then evaluated using the estimates from the two-segment empirical model. In other words, all six algorithms' performance results are based on the behavioral estimates from the two-segment empirical model; however, the one-segment algorithms' decisions are based on weighted averages of those estimates, whereas the two-segment algorithms' decisions are based on the heterogeneous estimates. Therefore, the two-segment algorithms should perform no worse than the one-segment algorithms.

Similarly, the discussion of goals in §5.2.1 implies that algorithms that allow the optimal importance of slot monetization and audience retention to vary throughout the break should do at least as well as algorithms that focus on either goal in isolation. However, it is not clear ex ante whether "sorting-only" algorithms will outperform the "traditional" algorithms. The performance of the traditional algorithms relative to sorting-only algorithms should increase with heterogeneity in the advertisers' willingness to pay; although both algorithms consider this factor,

---

[21] An aggregation of WTP parameters is not necessary, as our model does not differentiate advertisers' WTP across segments.

Table 6   Relative Algorithm Performance Measures

One-segment algorithms	Audience value Avg. (SD)	Network revenue Avg. (SD)	Two-segment algorithms	Audience value Avg. (SD)	Network revenue Avg. (SD)
Traditional sorting and selection	224 (13)	170 (10)	Traditional sorting and selection	225 (14)	171 (10)
Homogeneous sorting only	207 (18)	151 (9)	AVMA sorting only	212 (13)	165 (16)
Homogeneous sorting and DP	240 (17)	175 (11)	AVMA	252 (17)	188 (14)
Relative performance			Relative performance		
Homogeneous sorting only/ Traditional sorting and selection	93% (5)	89% (5)	AVMA sorting only/ Traditional sorting and selection	95% (6)	97% (7)
Homogeneous sorting and DP/ Traditional sorting and selection	107% (5)	103% (3)	AVMA/ Traditional sorting and selection	112% (6)	110% (6)
Homogeneous sorting and DP/ Homogeneous sorting only	117% (10)	116% (7)	AVMA/ AVMA sorting only	119% (7)	114% (11)
			AVMA/ Homogeneous sorting and DP	105% (7)	108% (6)

Note. Standard deviations are shown in parentheses.

traditional algorithms give it greater weight. On the other hand, the performance of traditional algorithms relative to sorting-only ones should decrease with heterogeneity in continuation probabilities, as these play no role in ranking ads in the traditional algorithms.

5.2.4.   Results. To compare the algorithms' performance, we drew 200 times from the joint distributions of the parameter estimates and ran all six algorithms on each set of draws. The top half of Table 6 reports the algorithms' performance statistics in terms of audience value and network revenues. The bottom half summarizes the distributions of ratios of selected pairs of algorithm performance statistics.

On average, AVMA produces 10% more audience value and 12% more network revenue than the traditional focus on slot monetization alone. The results show that allowing the optimal trade-off between slot monetization and audience retention to vary throughout the break is superior to focusing on either goal in isolation. We can also see, by comparing the first two rows of the table, that the slot monetization algorithms usually outperform the sorting-only algorithms. This was not true for every draw of parameter estimates, but it occurred for the majority of draws and happened more frequently for the one-segment algorithms than for the two-segment algorithms. It suggests that implementing an optimal advertising ordering heuristic without a dynamic programming component would likely reduce network revenues relative to their traditional practices.

Table 6 also compares algorithm performance across the number of segments considered, shown on the left and right sides of the table. The results

support the prediction that the two-segment algorithms would perform no worse than the one-segment algorithms.[22]

On average, AVMA produces 5% more audience value and 8% more revenue than a single-segment algorithm that optimally sorts and selects ads. The standard errors on these statistics are 7% and 6%, respectively, which may give the incorrect impression that there were cases in which the single-segment algorithm outperformed the multisegment algorithm. In fact, the distribution was asymmetric with a minimum at 100%: for all 200 sets of parameter draws, AVMA produced weakly higher audience value and network revenue than any of the other algorithms. The high standard deviations come from heavy right tails of the empirical distributions of performance ratios. In 28% of all scenarios, the homogeneous sorting and DP algorithm found the same selection and ordering of ads as AVMA. However, when this homogeneous sorting/DP algorithm failed to find the same selection and ordering of ads, network revenues were 9% lower on average, and sometimes as much as 30% lower.

Single-digit revenue increases may look modest, but it is important to remember that a small increase in revenues often implies a large increase in profits, and a small increase in a $70 billion business may

---

[22] The performance of the traditional sorting and selection algorithm was nearly identical in both one- and two-segment versions. The small differences between the two are because, as explained above, this algorithm orders advertisements randomly, so to smooth out those random fluctuations, it was run 1,000 times for the one-segment model and a separate 1,000 times for the two-segment model. The average of these 1,000 runs was taken as the algorithm's performance for each draw of the parameters.

be a substantial gain. We speculate that these results may actually understate the gains available from considering heterogeneity in selling TV advertisements. As discussed in §4.3, the audience and advertising data were not well suited to estimate heterogeneity in advertisers' willingness to pay for viewing segments. If there are quantifiable differences in segments' values to advertisers, there would be further gains available from valuing their retention differentially, as AVMA is designed to do.

5.3. AVMA Feasibility

AVMA performs well on both absolute and relative standards. Next, we consider whether it runs quickly enough to address the scale of a typical network's problem. AVMA's running time is $O(K^{S-1}Jnm^S)$, with $m = \log_{1-\delta}(\delta/J)$, where K is the resolution of the sorting rules considered and δ governs the precision of the dynamic programming subroutine. Although we do not expect AVMA to run in real time, we want to investigate whether there may be scales at which the algorithm is not a feasible means to solve the network's problem.

Determining whether AVMA scales to the typical network's problem requires defining the size of that problem. Some aspects are readily observable; the modal break is eight slots long, and $S = 2$ segments provides the best fit to the tuning data. The primary unknown is n, the number of potential advertisements.

Figure 4 shows how AVMA performance and computation time vary with n, for $J = 8$, $K = 15$,

$\delta = 0.07$, and considering both $S = 2$ and $S = 3$. As expected, the algorithm's computation time increases approximately linearly in the number of potential advertisements. Even with $n = 200$, AVMA runs in about 12 minutes on a single computer for the two-segment model. The results also show that audience value tends to increase with additional potential advertisements. This is what one would expect when the number of competing advertisements increases but the number of slots stays the same. However, past about $n = 115$, there is no apparent gain from including more advertisements in the auction. The three-segment model displays similar patterns.

Figure 5 shows how AVMA's performance and computation time vary with its two approximation parameters, holding n fixed at 100 and with $S = 2$. The first panel shows that computation time increases rapidly as we tighten the dynamic programming approximation parameter δ; however, maximal performance is already found at $\delta = 0.07$ with a trivial computation cost. The second panel shows that, as expected for $S = 2$, computation time is linear in K; the maximal audience value is found at a relatively low level of $K = 5$.

These computation-time figures may overstate the cost of running AVMA somewhat because they are based on serial processing on a single computer. Algorithm 2 in Appendix A can easily be parallelized so that computation time can be nearly divided by the number of cores available. It is straightforward to see how these figures could be combined with information about the cost of processing power to find the

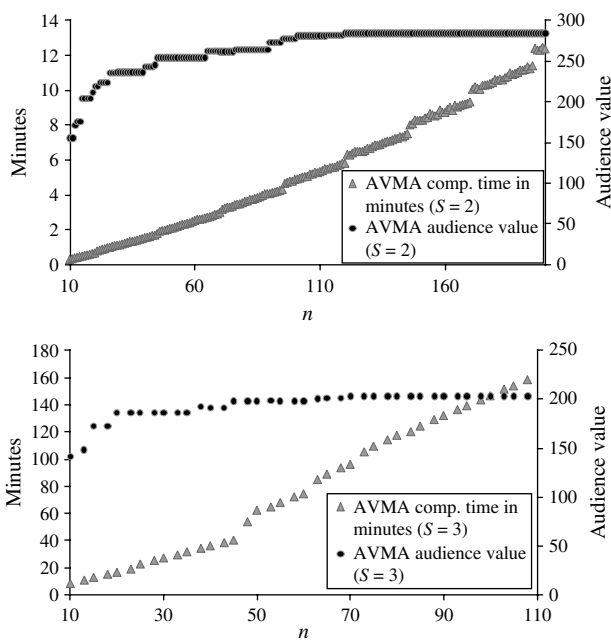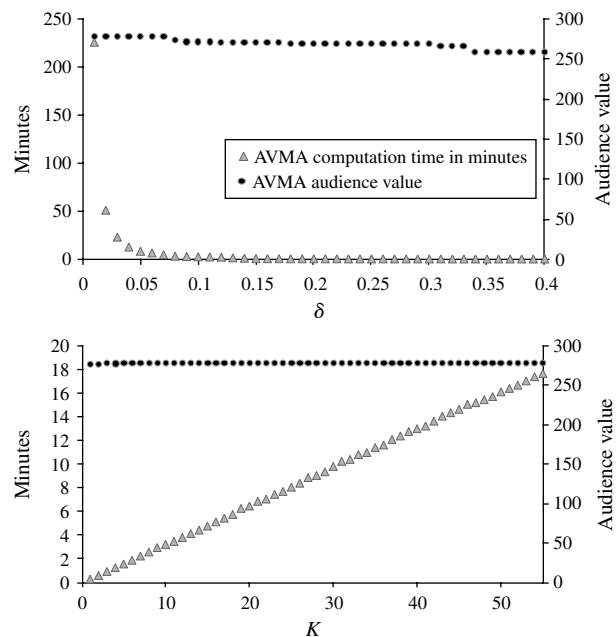**Figure 4**   **Computation Time and Audience Value as a Function of** n

Figure 5   **How Computation Time and Audience Value Change with** δ **and** K ($S = 2$)

optimal choices of δ and K for given n, J, and S. These results suggest that computational costs should be unlikely to dissuade a television network from adopting AVMA.

5.4. When Is AVMA *Un*necessary?

In the course of running many simulations, we have gained a good sense for the settings in which AVMA would be most profitably employed. There are also situations in which an algorithm as complex as AVMA may not be required.

There will be some settings in which the network's problem is simple, in that the number of slots (J) and potential advertisements (n) are relatively small. Such scenarios are more likely for niche-oriented networks or in countries that limit the amount of time that television networks may give to advertising. In those settings, the network could simply calculate the global optimum through brute force, rendering AVMA unnecessary.

In other settings, the network's problem may be very complex, but the difference among available solutions may be too small to justify a change in business model. For example, if ad-specific continuation rates did not differ much across potential advertisements, then little value could be gained from trying to eliminate audience externalities. Similarly, if continuation rates *do* vary across advertisements but covary positively with advertiser willingness to pay for audience, then the network's highest-value customers are also its best customers, in the sense that they keep the audience tuned to the channel. In yet another scenario, continuation rates may vary, but that variation may be very small relative to the variance in advertiser willingness to pay. In that case, audience externalities would be a second-order issue. However, although AVMA would not be needed in these examples, it may be advisable to change the network's business model from using quasi-random sorting to using some smarter algorithm, such as the one in Kempe and Mahdian (2008).

The potential returns to AVMA are greatest in two general settings. First, when the problem is very complex and the global optimum is too costly to compute, AVMA tends to find a good approximation of it in a reasonable time frame. Second, AVMA is particularly useful when the population of advertisements contains multiple clusters that exhibit positive within-cluster covariation in both willingness to pay and continuation rates for a particular viewing segment. These alternatives need not be exclusive, as a network could elect to use different algorithms for different breaks, depending on the characteristics of the problem at hand.

6. Discussion

The recent proliferation of digital set-top boxes has made it possible to replace television audience *estimates* with audience *measurements*. This article has considered how networks might use precise commercial audience measurements to address audience externalities. It developed the Audience Value Maximization Algorithm AVMA, a framework to address the television network's problem of selecting, ordering, and pricing advertisements in a commercial break of endogenous length with slot-specific tune-away rates and anticompetition constraints. The algorithm projects ads into a Euclidean space and uses dynamic programming to search for the best available selection and ordering of ads. Prices are set using a VCG mechanism to make the algorithm strategy-proof and to ensure that advertisers are charged directly for the reach and frequency they receive.

An empirical model of television advertising and viewing was estimated using market data. It found substantial heterogeneity in advertisements' tune-away rates and in advertisers' willingness to pay for viewers. It also showed that one segment of viewers zaps ads habitually according to the ad's slot, whereas the other does not. Both segments respond to advertising creative factors, albeit at different rates.

The parameter estimates were used in a series of simulations to gauge the proposed algorithm's performance. These simulations showed that AVMA regularly achieves the global optimum. They predict that, on average, network revenues are 10% higher under AVMA than the traditional sorting and selection procedure and 5% higher than an alternative that does not consider audience heterogeneity. It was also demonstrated that the algorithm runs quickly enough to implement at scale.

Although the predicted immediate returns are considerable, eliminating audience externalities could have even greater long-run consequences. Advertisers would have a direct financial incentive to choose commercial content to keep audiences watching, implying that ad elements that are attention-grabbing but annoying would become more expensive. This alignment of audience and advertiser incentives could potentially provide a more pleasant television-viewing experience and a reduction of audience losses. The resulting increase in viewers' utility of live television might even reduce viewers' ad avoidance habits and time-shifting behavior. It also might enable television networks to retain a greater share of channel margins by disintermediating existing resellers of their advertising inventory.

AVMA adoption may be accelerated by peer effects between networks: as one network eliminates or reduces audience externalities, other networks are

likely to feel increased pressure to do the same. Suppose that Network 1 starts charging for audience loss while Network 2 does not. As a result, slots on Network 1 will become relatively cheaper for ads, causing low switching and relatively more expensive for ads causing high zapping. Advertisers will sort themselves accordingly: some high-zapping advertisements will shift from Network 1 to Network 2 in equilibrium, whereas some low-zapping ads will shift from Network 2 to Network 1. This sorting effect will increase Network 2's audience losses and, consequently, its potential gain from eliminating audience externalities as well.

The algorithms proposed here may also find applications beyond television. The modeling framework could potentially represent any advertising-supported medium that is broadcast to heterogeneous, ad-averse audience segments. Such media might include commercial radio, online video, or display advertising.

There are a number of implementation details of the proposed algorithms that are likely to vary across television networks or countries. Furthermore, there are concerns about how AVMA could be implemented to conform to current business practices. Online Appendix C discusses how the proposed framework could be extended to address many of these issues, including public policy implications, order effects among commercials, multiple commercial breaks, advance selling, audience bundling, and additional constraints, as well as predicting continuation probabilities for new advertisements. Some of these topics may present exciting opportunities for future research.

More generally, we are hopeful that the television industry will continue current trends toward standardizing and using the newly available set-top box data on audience behavior. To some degree, correcting audience externalities is low-hanging fruit. These data may also help content providers optimize in other areas such as program selection, audience bundling, and product placement.

Supplemental Material

Acknowledgments

Appendix A. Formal Specification of AVMA

Here, we provide more formal specifications of the algorithms used to select and order ads in AVMA. We first describe the projection algorithm that is used to find the best ordering of selected ads.

Algorithm 1 (Projection algorithm)
1: **for** a representative set of vectors $\boldsymbol{\alpha} = (\alpha_s)_{s=1}^{S-1} \in [0, \pi/2]^{S-1}$ **do**
2:   **for** $s = 1, \ldots, S-1$ **do**
3:     Let $x_s = \cos(\alpha_s) \cdot \prod_{j<s} \sin(\alpha_j)$.
4:   Let $x_S = \prod_{j=1}^{S-1} \sin(\alpha_j)$.
5:   **for all** ads $a = 1, \ldots, n$ **do**
6:     Let $t_a = \sum_{s=1}^{S} x_s \nu_s b_a^{(s)} c_a^{(s)} / (1 - c_a^{(s)})$.
7:   Sort the ads by nonincreasing t_a values.
8:   **if** $J = n$ (all ads can be shown) **then**
9:     Consider the sorted order.
10:   **else**
11:     Use dynamic programming (Algorithm 2 below) to select J ads, and consider the sorted order.
12:   Use Equation (2) to compute the aggregate welfare from the selected ads and order.
13:   If the current selection and order is better than the previous best, update the estimate of best.

The first step of Algorithm 1 specifies a set of lines, representing sorting rules to be considered. There are many alternative ways to do this. Systematic exploration ensures the best possible coverage of the space and has consistently found globally optimal solutions in our experiments, so it is the option used in our subsequent implementations. We use $S-1$ nested loops over the possible angles from 0 to $\pi/2$. The values explored for α_s are multiples of $\pi/2k_s$, where $k_s = 1 + \lceil K \cdot \prod_{j<s} \sin(\alpha_j) \rceil$ is the number of values explored, and K a parameter that determines the resolution. Other possibilities include sampling vectors from the hypersphere randomly, or obtaining pseudo-random collections of points by using $(S-1)$-dimensional Halton sequences (Halton and Smith 1964).

The dynamic programming algorithm used as a subroutine takes an advertisement sorting from Algorithm 1 as an input. Algorithm 2's accuracy is governed by a parameter δ. The closer δ is to zero, the greater the accuracy, but the greater the cost in terms of computation time and storage. A preprocessing step rounds all continuation rates $c_a^{(s)}$ and all audience fractions ν_s down to the closest power of $(1-\delta)$; values smaller than δ/J are rounded down to 0. There are $m = \log_{1-\delta}(\delta/J)$ different powers of $(1-\delta)$ to which values could be rounded.

The key to the algorithm is to maintain a dynamic programming table H, which has $S+2$ dimensions. It contains, for each tuple $\langle a, j, P_j^{(1)}, \ldots, P_j^{(S)} \rangle$, the maximum welfare that can be obtained using ads $a, \ldots, n$ (without changing

their order) in the slots $j, \ldots, J$, if the remaining fractions of viewers in slot j for the segments are $P_j^{(1)}, \ldots, P_j^{(S)}$. A second table R tracks the inclusion of ads in optimum subsolutions.

The important part is the dynamic programming loop, which computes the optimum solution for each subproblem by either assigning ad a to slot j or not assigning it, depending on which gives the better solution with the remaining slots and ads. An important operational note is that the best way to store R and H is to index them with $\log_{1-\delta} P_j^{(s)}$ (which takes on values from $0, 1, \ldots, m$ with a special value ∞ when $P_j^{(s)} = 0$), not the actual $P_j^{(s)}$ value.

The running time of Algorithm 2 is $O(Jnm^S)$. It thus grows exponentially in S. Since m grows as δ becomes smaller, δ must be chosen to balance accuracy with computation time. Section 5.3 showed that this trade-off can be made in favor of accuracy, as the algorithm exhibits reasonable running time for problems of realistic scale.

Algorithm 2 (Dynamic programming algorithm)
1: {Preprocessing of probabilities}
2: **for all** ads a and segments s **do**
3:    **if** $c_a^{(s)} < \delta/J$ **then**
4:       Round $c_a^{(s)}$ down to 0.
5:    **else**
6:       Round $c_a^{(s)}$ down to the nearest power of $1 - \delta$.
7: **for all** segments $s = 1, \ldots, S$ **do**
8:    **if** $\nu_s < \delta/J$ **then**
9:       Round ν_s down to 0.
10:    **else**
11:       Round ν_s down to the nearest power of $1 - \delta$.
12: {Initialization of DP array}
13: **for all** $a, j, q_1, \ldots, q_S$ **do**
14:    Let $H[a, J+1, q_1, \ldots, q_S] = 0$.
15:    Let $H[n+1, j, q_1, \ldots, q_S] = 0$.
16: {Central part of DP loops}
17: **for** $j = J$ downto 1 **do**
18:    **for** $a = n$ downto 1 **do**
19:       **for all** $q_1, \ldots, q_S = 1, 1-\delta, \ldots, (1-\delta)^{m(J+1-j)}, 0$ **do**
20:          **if** $H[a+1, j, q_1, \ldots, q_S] > H[a+1, j+1, q_1 c_a^{(1)}, \ldots, q_S c_a^{(S)}] + \sum_s b_a^{(s)} c_a^{(s)} q_s$ **then**
21:             Let $H[a, j, q_1, \ldots, q_S] = H[a+1, j, q_1, \ldots, q_S]$ and $R[a, j, q_1, \ldots, q_S] = 0$.
22:          **else**
23:             Let $H[a, j, q_1, \ldots, q_S] = H[a+1, j+1, q_1 c_a^{(1)}, \ldots, q_S c_a^{(S)}] + \sum_s b_a^{(s)} c_a^{(s)} q_s$ and $R[a, j, q_1, \ldots, q_S] = 1$.
24: {Postprocessing to reconstruct the solution}
25: Let $j = 1, a = 1$, and $q_s = \nu_s$ for each $s = 1, \ldots, S$.
26: **While** $j \leq J$ **do**
27:    **if** $R[a, j, q_1, \ldots, q_S] = 0$ **then**
28:       Let $a = a + 1$.
29:    **else**
30:       Output "Include ad a in position j."
31:       **for all** $s = 1, \ldots, S$ **do**
32:          Let $q_s = q_s \cdot c_a^{(s)}$.
33:       Let $j = j + 1$ and $a = a + 1$.

Appendix B. Empirical Viewing Model

This appendix specifies the model and estimation strategy used to estimate advertising creative-specific continuation probabilities. The modeling paradigm is developed based on the approach of Chintagunta and Gupta (1994), which specified discrete heterogeneity within a logit framework and used consumer characteristics to assign consumers to segments. The model is used for the dual purposes of learning about drivers of advertising avoidance and producing estimates that allow for a realistic comparison of AVMA to alternative algorithms.

The logit model assumes that each time an ad is shown on a consumer's screen, he has the opportunity to continue that advertising exposure or to stop it (whether by switching channels or turning off the television). An alternative approach would be to use a hazards framework based on individual STB-level advertising avoidance decisions. While this approach seems promising, we do not adopt it for three reasons. The primary reason is that a considerable body of marketing literature (e.g., Goettler and Shachar 2001, Rust and Alpert 1984, Rust et al. 1992, Shachar and Emerson 2000) has used discrete choice models to estimate models of television viewing behavior. For results differing from previous authors' findings, working within a similar paradigm eliminates a major possible confound as an explanation for those differences. Second, privacy laws suggest that industry operators will likely be working with aggregate data. For example, in the United States, the Cable Communications Policy Act of 1984 bans cable television operators from sharing individual-level viewing data with any third parties (e.g., Nielsen) without the subscriber's explicit consent. However, television operators are allowed to share aggregate data, and the discrete choice model is more easily adapted to an aggregate data setting. Third, even if the law did allow for analysis of household-level viewing data, estimation speed would be a first-order concern that would likely (but not certainly) favor the discrete-choice model.

Formalizing the model, set-top box h in segment s exposed to any part of ad a during slot j of break k obtains utility

$$u_{h,a,j,k} = \alpha_{a,s} + \beta_{j,s} + \gamma_{k,s} + X_{h,a,j,k}\delta_s + \epsilon_{h,a,j,k}. \quad (B1)$$

$\alpha_{a,s}$ captures the intrinsic utility obtained by a viewer in segment s who watches ad a, and $\beta_{j,s}$ is the utility of a viewer in segment s watching slot j; the first slot dummy is excluded for identification. $\gamma_{k,s}$ is the utility of a viewer in segment s watching break k, defined only for the first and last breaks within each program due to data sparseness.

$X_{h,a,j,k}$ captures all observed data that may influence switching. We draw from earlier work on set-top box viewing data (e.g., Danaher 1995; Schweidel and Kent 2010, 2011; Siddarth and Chattopadhyay 1997) in constructing this set. These include the number of times that box h has been exposed to ad a in the previous seven days. Seven days was found to be more highly correlated with tune-away than alternatives 3, 5, 10, and 14. Another control is the number of minutes the box has been watching the channel. $X_{h,a,j,k}$ also includes dummy variables for weekday, half-hour, and program genre. δ_s measures how the variables in $X_{h,a,j,k}$ influence the continuation probability of a viewer in segment s.

As is standard in many discrete choice models, $\epsilon_{h,a,j,k}$ is an idiosyncratic preference shock distributed i.i.d. Type 1

extreme value with scale parameter 1, and the value of tune-away is normalized to zero for identification. In the following, $d_{h,a,j,k}$ is an indicator variable that equals 1 if box h was observed to continue watching until the end of ad a in slot j of break k. Conditional on being in segment s, the probability of this event is

$$\text{Prob}[d_{h,a,j,k} = 1 \mid h \in s]$$
$$= \frac{\exp(\alpha_{a,s} + \beta_{j,s} + \gamma_{k,s} + X_{h,a,j,k}\delta_s)}{1 + \exp(\alpha_{a,s} + \beta_{j,s} + \gamma_{k,s} + X_{h,a,j,k}\delta_s)}. \quad \text{(B2)}$$

The probability that box h is in segment $s \in \{1, \dots, S\}$ is

$$\nu_{hs} = \exp(Z_h \lambda_s) \Big/ \Big(1 + \sum_{s'=1}^{S-1} \exp(Z_h \lambda_{s'})\Big), \quad \text{(B3)}$$

where Z_h is a vector containing two usographics (television viewing propensity and ad tune-away propensity) corresponding to STB h, and λ_s captures the effect of usographics on the probability that STB h is in segment s. Both elements of λ_S are normalized to zero for identification.

The unconditional probability that box h watches the end of ad a in slot j in break k is

$$\text{Prob}[d_{h,a,j,k} = 1] = \sum_s \nu_{hs} \text{Prob}[d_{h,a,j,k} = 1 \mid h \in s]. \quad \text{(B4)}$$

The likelihood contribution of box h is

$$L_h = \sum_s \nu_{hs} \prod_{a,j,k} (\text{Prob}[d_{h,a,j,k} = 1 \mid h \in s])^{d_{h,a,j,k}}$$
$$\cdot (\text{Prob}[d_{h,a,j,k} = 0 \mid h \in s])^{1 - d_{h,a,j,k}}, \quad \text{(B5)}$$

and the total likelihood is the product of (B5) across all STBs.

After estimating the model, we use the parameter estimates to construct the estimated continuation rate of each ad a in segment s as

$$\hat{c}_a^{(s)} = \frac{1}{H_{ak}} \cdot \sum_{h,k} \nu_{hs} \text{Prob}[d_{h,a,j,k} = 1 \mid h \in s], \quad \text{(B6)}$$

where H_{ak} is the number of boxes exposed to ad a in break k, and ν_{hs} and $\text{Prob}[d_{h,a,j,k} = 1 \mid h \in s]$ are calculated based on the parameter estimates and the posterior estimates of households' segment membership. The $\hat{c}_a^{(s)}$ parameters are used as inputs in the simulation evaluation in §5.

References

Abrams Z, Schwarz M (2008) Ad auction design and user experience. *Appl. Econom. Res. Bull.* 2:98–105.

Aggarwal G, Feldman J, Muthukrishnan SM, Pál M (2008) Sponsored search auctions with Markovian users. *Proc. 4th Workshop Internet Network Econom. (WINE)*, 621–628.

Athey S, Ellison G (2011) Position auctions with consumer search. *Quart. J. Econom.* 126(3):1213–1270.

Bronnenberg BJ, Dubé J-P, Mela CF (2010) Do digital video recorders influence consumers' brand purchases? *J. Marketing Res.* 47(6):998–1010.

Brown TJ, Rothschild ML (1993) Reassessing the impact of television advertising clutter. *J. Consumer Res.* 20(1):138–146.

Burke RR, Srull TK (1988) Competitive interference and consumer memory for advertising. *J. Consumer Res.* 15(1):55–68.

Chintagunta P, Gupta S (1994) On using demographic variables to determine segment membership in logit mixture models. *J. Marketing Res.* 31(1):128–136.

Clarke EH (1971) Multipart pricing of public goods. *Public Choice* 11(Fall):17–33.

Crampton P, Shoham Y, Steinberg R, eds. (2006) *Combinatorial Auctions* (MIT Press, Cambridge, MA).

Danaher PJ (1995) What happens to television ratings during commercial breaks? *J. Advertising Res.* 35(1):37–47.

Danaher PJ, Mawhinney DF (2001) Optimizing television program schedules using choice modeling. *J. Marketing Res.* 38(3):298–312.

Downey K (2006) Study: First ad does get more eyeballs. *Media Life Magazine* (August 21) http://www.medialifemagazine.com/study-first-ad-does-get-more-eyeballs/.

Edelman B, Ostrovsky M, Schwarz M (2007) Internet advertising and the generalized second price auction: Selling billions of dollars worth of keywords. *Amer. Econom. Rev.* 97(1):242–259.

Ferencz J, Teixeira T (2010) How much does advertising really cost? Working paper, Harvard University, Cambridge, MA.

Gaither J (2011) Is there an alternative ratings source in our future? *Richards Group* (blog), February 2011, http://blog.richards.com/richards-connections-february-2011.

Ghosh A, Mahdian M (2008) Externalities in online advertising. *17th Internat. World Wide Web Conf.*, 161–168.

Goettler RL, Shachar R (2001) Spatial competition in the network television industry. *RAND J. Econom.* 32(4):624–656.

Goldfarb A, Tucker C (2011a) Online display advertising: Targeting and obtrusiveness. *Marketing Sci.* 30(3):378–404.

Goldfarb A, Tucker C (2011b) Privacy regulation and online advertising. *Management Sci.* 57(1):57–71.

Gomes R, Immorlica N, Markakis E (2009) Externalities in keyword auctions: An empirical and theoretical assessment. *Proc. 5th Workshop Internet Network Econom. (WINE)*, 172–183.

Groves T (1973) Incentives in teams. *Econometrica* 41(4):617–631.

Halton JH, Smith GB (1964) Algorithm 247: Radical-inverse quasi-random point sequence. *Comm. ACM* 7(12):701–702.

Håstad J (1999) Clique is hard to approximate within $n^{1-\varepsilon}$. *Acta Mathematica* 182:105–142.

Jehiel P, Moldovanu B, Stacchetti E (1996) How (not) to sell nuclear weapons. *Amer. Econom. Rev.* 86(4):814–829.

Karp R (1972) Reducibility among combinatorial problems. Miller R, Thatcher J, eds. *Complexity of Computer Computations* (Plenum Press, New York), 85–103.

Kempe D, Mahdian M (2008) A cascade model for externalities in sponsored search. *Proc. 4th Workshop Internet Network Econom. (WINE)*, 585–596.

Kirkpatrick S, Gelatt CD, Vecchi MP (1983) Optimization by simulated annealing. *Science* 220(4598):671–680.

Kumar V, Schechner S (2009) Cable industry weighs its approach to targeted ads. *Wall Street Journal* (May 27) B6.

Mandese J (2004) Equitable's no longer questionable: Data reveals nets position some advertisers better than others. *MediaPost's MediaDailyNews* (September 17) http://www.mediapost.com/publications/article/18344/#axzz2cXZPMbvG.

Morgan D (2013) Want to shift ad dollars out of TV into Web video? Good luck with that. *AdAge* (March 15) http://adage.com/article/digitalnext/shift-ad-dollars-tv-web-video-good-luck/240393/.

Myerson R (1981) Optimal auction design. *Math. Oper. Res.* 6(1):58–73.

Rassenti SJ, Smith VL, Bulfin RL (1982) A combinatorial auction mechanism for airport time slot allocation. *Bell J. Econom.* 13(2):402–417.

Rust RT, Alpert MI (1984) An audience flow model of television viewing choice. *Marketing Sci.* 3(2):113–124.

Rust RT, Kamakura WA, Alpert MI (1992) Viewer preference segmentation and viewing choice models for network television. *J. Advertising* 21(1):1–8.

Schweidel DA, Kent RJ (2010) Predictors of the gap between program audiences and commercial audiences: An investigation using live tuning data. *J. Marketing* 74(3):18–33.

Schweidel DA, Kent RJ (2011) Introducing the Ad ECG: How the set-top box tracks. *J. Advertising Res.* 51(4):586–593.

Scott S (2010) A modern Bayesian look at the multi-armed bandit. *Appl. Stochastic Models Bus. Indust.* 26(6):639–658.

Shachar R, Emerson JW (2000) Cast demographics, unobserved segments, and heterogeneous switching costs in a television viewing choice model. *Marketing Sci.* 37(2):173–186.

Siddarth S, Chattopadhyay A (1997) To zap or not to zap: A study of the determinants of channel switching during commercials. *Marketing Sci.* 17(2):124–138.

Stourm V, Bax E (2013) Charging for ad impact on viewer experience: Marketplace dynamics and Pareto efficiency. Working paper, University of Pennsylvania, Philadelphia.

Teixeira TS, Wedel M, Pieters R (2010) Moment-to-moment optimal branding in TV commercials: Preventing avoidance by pulsing. *Marketing Sci.* 29(5):783–804.

Television Bureau of Advertising (TVB) (2013) TV basics. Report, Television Bureau of Advertising, New York. http://www.tvb.org/media/file/TV_Basics.pdf.

Varian H (2007) Position auctions. *Internat. J. Indust. Organ.* 25(6):1163–1178.

Vickrey W (1961) Counterspeculation, auctions, and competitive sealed tenders. *J. Finance* 16(1):8–37.

Webb PH, Ray ML (1979) Effects of TV clutter. *J. Advertising Res.* 19(3):7–12.

Wilbur KC (2008) A two-sided, empirical model of television advertising and viewing markets. *Marketing Sci.* 27(3):356–378.

Wilbur KC (2013) Advertising avoidance in live television. Working paper, Duke University, Durham, NC.

Wilbur KC, Goeree MS, Ridder G (2008) The effects of advertising and product placement on television audiences. Working paper, University of California, San Diego, San Diego. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1151507.

Woltman Elpers JLCM, Wedel M, Pieters RGM (2003) Why do consumers stop viewing television commercials? Two experiments on the influence of moment-to-moment entertainment and information value. *J. Marketing Res.* 40(4):437–453.

Wortman J, Vorobeychik Y, Li L, Langford J (2007) Maintaining equilibria during exploration in sponsored search auctions. *Proc. 3rd Workshop Internet Network Econom.* (WINE), 119–130.

Zigmond D, Interian Y, Lanning S, Hawkins J, Mirisola R, Rowe S, Volovich Y (2009) When viewers control the schedule: Measuring the impact of digital video recording on TV viewership. Paper presented at the Key Issues Forums at ARF Audience Measurement Conference, June 23, Advertising Research Foundation, New York.