What can Television Networks Learn from Search Engines?
How to Select, Price, and Order Ads to Maximize Advertiser Welfare

David Kempe\textsuperscript{1} and Kenneth C. Wilbur\textsuperscript{2}

July 3, 2009

Abstract: Television advertising revenues are falling after years of growth due, in part, to digital video recorder proliferation. We consider the television network’s problem of how optimally to select, price, and order advertisements in order to maximize audience value. We propose that television networks should shift from managing and selling \textit{time} to managing and selling \textit{viewer attention} in order to eliminate negative externalities among advertisers within a break. We show that there is no optimal advertisement ordering heuristic when the viewing audience contains multiple segments. We propose the Audience Value Maximization Algorithm (AVMA), a computationally feasible second-best solution that considers many possible advertisement orderings within a dynamic programming framework. AVMA can be extended to accommodate advance selling, audience bundling, and branded entertainment.

\textsuperscript{1}Assistant Professor, Computer Science, Viterbi School of Engineering, University of Southern California. dkempe@usc.edu, http://www.rcf.usc.edu/~dkempe/

\textsuperscript{2}Assistant Professor, Fuqua School of Business, Duke University. kennethwilbur@gmail.com, http://kennethwilbur.com
Television viewers are increasingly avoiding advertising, advertisers are able to observe this ad-avoidance, and television advertising revenue growth is suffering.

Viewers employ two easily measurable strategies for skipping ads: fast-forwarding and channel switching. 31% of US households had Digital Video Recorders (DVRs) as of May 2009 (TVB 2009). Bronnenberg, Dube, and Mela (2009) analyze DVR usage data and report that viewers fast-forward through 71% of ads in recorded programming. In addition, digital cable has greatly expanded the number of networks available, increasing viewers’ utility and motivation to switch channels to avoid advertisements.

At the same time, advertisers are increasingly able to observe viewers’ ad avoidance and can obtain precise advertisement avoidance data. Second-by-second DVR usage data is commercially available from TiVo. Cable and satellite companies’ digital set-top boxes can record moment-by-moment viewing actions in millions of households. Since September 2007, advertising deals have been based on programs’ average commercial minute ratings (“C3”), but many analysts expect more granular audience data in the future.

Figure 1 presents TNS Media Intelligence data showing that network television revenues have started to dip after a long expansion. Spot TV revenues fell sharply in 2009. Cable television revenues were down year-over-year for the first time in recent memory.

Figure 1. Trends in Television Advertising Revenues ($B)
Advertiser revenue commitments in the annual upfront market are expected to drop between 10-20% in 2009 (Steinberg 2009). This pattern stands in contrast to continued annual increases in time spent viewing television, which was up 2% to 153:27 hours per month in 2009Q1 (Friedman 2009).¹

Historically, television networks have managed and sold advertising time. We propose a fundamental shift: networks should manage and sell the truly scarce resource in this industry—viewers’ attention. We present the problem of selecting, ordering, and pricing ads in order to maximize network profits, and show how television networks can adapt recently developed algorithms for online ad placement. We first prove that when multiple viewing segments exist, scheduling ads via simple heuristics can perform very badly. We then propose the Audience Value Maximization Algorithm (AVMA), an intuitive algorithmic solution that uses advertisers’ willingness to pay and viewers’ propensity to switch in order to select, order, and price ads optimally within a commercial break.

While our motivating examples and data come from the US market, the problems and solutions we present will likely apply in many other countries’ television markets.

In the next section we describe the negative externalities among advertisers present under current business practices. Section 2 discusses the previous literature to which we contribute. Section 3 introduces a baseline model of optimal ad ordering given homogeneous consumers. Section 4 both establishes that simple heuristics can perform very badly when we add consumer segments to the model, and introduces the AVMA. Section 5 concludes with a discussion of how to extend the algorithm to allow for advance selling, audience bundling, branded entertainment, practical challenges to implementing the algorithm, and possible extensions.

1. Negative Externalities within a Commercial Break.

Television networks have traditionally scheduled television commercials within commercial breaks on an “equitable” basis (Mandese 2004, Downey 2006). Networks are understood to assign ad slots randomly across advertisers. The first position within a commercial break is known to be the most valuable since it is well known that some viewers switch away during the break and do not switch back right away. However, this value is almost never priced.³ The last position in the break is also thought to be of greater value than intermediate positions because
some viewers are switching back at that point. The last position is often retained for promotions of upcoming network programs.

Under the current pricing structure, advertisers have limited incentive to retain viewers. Certainly they have some incentive—an ad that causes all viewers to switch channels may have little positive effect on sales and may engender ill will toward the advertiser’s brand. But imagine a stimulus placed in an ad that would simultaneously increase advertising effectiveness by 20% while causing a 5% audience loss (e.g., a car dealer screaming at the camera). Because the 1.2 effectiveness multiplier outweighs the 0.05 audience reduction, the profit-maximizing advertiser will include this stimulus in its ad. But this action results in a negative externality: it reduces the audience remaining to watch commercials in the following ad slots.

If every advertiser were to make similar choices, the number of viewers watching commercials would be reduced substantially. However, if advertisers were penalized or rewarded for their ads’ audience losses or gains, they would design ads to hold viewer attention to a greater degree, enhancing overall efficiency. This is the primary strategic value of the Audience Value Maximization Algorithm that we introduce in this paper.

This core idea is not entirely new. Many internet media firms use measures of consumer acceptance to sell and place advertisements. Google introduced the use of click-through rates into the auction mechanism it uses to sell keywords in 2002 (Battelle 2005). Advertisers with high rates of user acceptance get more desirable ad positions and pay lower costs, other factors being equal. The use of viewer acceptance increases viewer utility of ads and reduces ad avoidance. It also increases search engine revenues by increasing the number of paid clicks and benefits those advertisers whose ads are most useful to viewers by increasing traffic and lowering advertising costs. Television networks will realize similar benefits if they adopt the AVMA we propose here.

2. Relevant literature on ad avoidance
There is a literature in marketing and advertising that documents the strategies television viewers use to avoid advertisements. The strategies most likely to affect network revenues are changing channels (“zapping”) and fast-forwarding (“zipping”) since these actions can be measured cost-effectively. Danaher (1989) investigated Peoplemeter data from New Zealand and found that network ratings fell by a net 5% during commercial breaks due to a 10% audience loss to
zapping and a 5% audience gain from viewers leaving other networks. However, the context of this study was a 3-channel environment in which simultaneous ad breaks were commonplace. Van Meurs (1998) looked at Peoplemeter data from the Netherlands and found a 21.5% net audience loss during commercial breaks in a 21-channel environment. These two results suggest, intuitively, that zapping becomes more common as viewers have more channels available.

There is also research on viewer fast-forwarding to avoid commercials. Cronin and Menelly (1992) designed two studies to measure the zipping rate among viewers equipped with VCRs. Among college students, zipping affected 75% of all ad exposures; in a more representative sample, it affected 62% of ad exposures. Downey (2007) reported that, based on a large sample of TiVo users, broadcast network commercials lose an average of 59% of the program’s audience to fast-forwarding, but the first ad in the break lost only 49% on average. Commercials on the 10 largest cable networks lost 42% of the program’s rating on average, but the first commercial in the break lost an average of only 29%.5

A few studies have related advertisement avoidance to advertising content. Woltman Elpers, et al. (2003) found that viewers chose to stop viewing commercials with low entertainment content. Teixeira, Wedel, and Pieters (2008) used eye-tracking technology and found that focal deviations from main elements in the ad storyline tended to increase ad avoidance. Wilbur, Goeree, and Ridder (2008) estimated audience reactions to specific ads in a sample of 10,000 prime time broadcast network program-hours from 2004-2007. They found that viewer utility of advertising was driven primarily by product category and ad content. Interian et al. (2009) use set-top box data and report substantial heterogeneity in the rate at which households tune away from various advertisements.

We are aware of just one other paper that proposes an algorithm to select and order advertisements. Kimms and Muller-Bungart (2007) introduce a revenue management algorithm to select and order ads from a set of potential advertisements. However, they do not model the responsiveness of audience size to advertiser identity, the central source of complication in this paper. Accordingly, they also do not model advertiser demand as a function of audience size.

3. A simple model of ad ordering
In this section we simplify the problem to convey the basic intuition behind our proposed algorithms. The network’s task is to choose from a set of $a = 1...n$ potential advertisements in
order to fill \( j = 1 \ldots J \) slots within a commercial break. Slots are numbered chronologically according to the order in which they appear. We denote the advertisement \( a \) shown in slot \( j \) with \( a_j \). We assume viewers are homogeneous. We normalize the measure of initial audience size to 1 without loss of generality.

Each ad has an attention rate \( c_a \), the ratio of audience retained at the end of an ad. \( c_a \) has two possible interpretations. If no viewers join the audience during the ad, it follows that for every 100 viewers who were watching at the beginning of ad \( a \), \( 100 \times c_a \) will watch the entire ad. If the viewer stops watching the commercial, we assume she watches another channel or fast-forwards using a digital video recorder and subsequently misses all of the ads remaining in the commercial break. This first interpretation implies that \( c_a \) is bounded above by 1 and represents the gross audience loss induced by the ad.

The second interpretation allows for viewer inflow during the commercial, as viewers may be turning on their television sets or leaving other channels. If we assume a constant proportional inflow rate \( x \), and if we assume that the ad causes a proportion \( y \) of all viewers to leave the audience, then for every 100 viewers who were watching at the beginning of ad \( a \), \( 100 \times (1 + x - y) \) will be watching at the end. We would then define the continuation probability as \( c_a = (1 + x - y) \). Advertisers’ continuation probabilities may differ according to many factors, including the relevance of the product to the viewing audience, the entertainment value of the commercial, or the match between the ad and the program.

The commercial \( a \) in slot \( j \) has an expected audience of

\[
P_j = \prod_{i=1}^{j} c_{a_i}.
\]

Equation (1) makes it clear that the audience watching slot \( j \) depends on the attention rates of the commercials placed in the previous slots.

Each advertiser \( a \) has a privately held willingness to pay \( v_a \) for one viewer to be exposed to any part of its ad. The advertiser indicates its willingness to pay for the first advertising slot with a bid \( b_a \). For now we assume that advertisers bid truthfully and then show this is incentive compatible in the auction described below.
The network’s objective is to select and order ads to maximize its advertising revenue. If advertisers bid truthfully (we show below that they do), then total advertiser welfare is

$$W = \sum_{i=1}^{J} b_i \prod_{j=1}^{i} c_{a_j}.$$  \hspace{1cm} (2)

3.1. Ordering Ads with Homogeneous Viewers

Kempe and Mahdian (2008) show that under the preceding conditions the advertisement ordering that maximizes total advertiser welfare satisfies

$$\frac{b_{a_1} c_{a_1}}{1 - c_{a_1}} \geq \frac{b_{a_2} c_{a_2}}{1 - c_{a_2}} \geq \ldots \geq \frac{b_{a_J} c_{a_J}}{1 - c_{a_J}}.$$  \hspace{1cm} (3)

The proof begins with a pairwise analysis. Suppose we evaluate two consecutive slots $$k$$ and $$k + 1$$. The advertiser welfare generated by those two slots is

$$W_{k,k+1} = b_k \prod_{j=1}^{k} c_{a_j} + b_{k+1} \prod_{j=1}^{k+1} c_{a_j} = \prod_{j=1}^{k+1} c_{a_j} (b_k c_{a_k} + b_{k+1} c_{a_{k+1}} c_{a_{k+1}}).$$  \hspace{1cm} (4)

If we reverse the ordering of the ads in $$k$$ and $$k + 1$$, welfare rises if and only if

$$b_k c_{a_k} + b_{k+1} c_{a_{k+1}} c_{a_{k+1}} > b_{k+1} c_{a_{k+1}} + b_k c_{a_k} c_{a_{k+1}} \Leftrightarrow \frac{b_k c_{a_k}}{1 - c_{a_k}} \geq \frac{b_{k+1} c_{a_{k+1}}}{1 - c_{a_{k+1}}}. \hspace{1cm} (5)$$

Starting from an arbitrary advertisement ordering, we can apply this pairwise analysis repeatedly to the $$J - 1$$ pairs of consecutive ad slots until it is no longer possible to rearrange any two ads $$k$$ and $$k + 1$$ to increase welfare. In the optimal ordering, it must be the case that (3) holds both for consecutive and nonconsecutive pairs of advertisements, yielding a simple heuristic for ordering commercials.

3.2. Selecting Ads with Homogeneous Viewers

While (3) tells us how to order a set of ads optimally, it says nothing about how to select them. It is not always optimal to select simply the top $$k$$ ads by this sorting. For instance, if the top $$J$$ ads had bids equal to 0.1 but $$c_a = 1$$, they would precede an ad $$a'$$ with a bid of 10 but $$c_{a'} = 0.9$$. For finite $$J$$, the network would do better if it replaced the $$J$$th ad with ad $$a'$$.

We can use a dynamic programming algorithm to choose the optimal set of $$k$$ ads. We first order the ads according to (3). Then we construct a table $$A_{n,J}$$ whose entry $$A[a,i]$$ contains the optimum value that can be obtained from a subset of ads $$a,...,n$$ in positions $$i,...,J$$. 

6
conditional on both ad \( a \) being placed in slot \( i \), and the expected audience of slot \( i \). Each entry in the table is given by the following recurrence:

\[
A[a, i] = \max(A[a + 1, i], b_a + c_a A[a + 1, i + 1] ) .
\]

(6)

Once the table is filled out, the solution is given in cell \( A[1,1] \).

The intuition here is as follows. The table is populated beginning with the bottom-right-corner. The bottom-right cell of the table, \( A[n, J] \), contains the revenue generated by placing the last ad in the last slot. The cell above, \( A[n − 1, J] \), shows the maximal revenue available from considering whether to place the second-to-last ad in the last slot. This maximal revenue is given by (6): it is either given by the revenue produced by putting the last ad in that last slot (\( A[n, J] \)), or by putting the second-to-last ad in that last slot (\( b_{n−1} \)). Iterating backwards, the table is then filled out recursively until arriving at the solution in \( A[1,1] \).

3.3. Pricing Ads with Homogeneous Viewers

It remains to be shown that an auction exists that induces truth-telling bids. Advertiser payments are computed in the following way: for any given set of ads \( A \), the optimal ordering is calculated according to (3), which implies an aggregate advertiser surplus \( W(A) \). The total payment of each included ad \( a \) is \( W(A) − W(A \setminus a) \), where \( W(A \setminus a) \) is the aggregate surplus produced by the optimal ranking when \( a \) is excluded from the set of potential advertisements. Each advertiser is only included if this surplus \( W(A) − W(A \setminus a) \) is less than its bid. Advertisers are then ranked by the amount of surplus produced, and each advertiser is charged the next-lowest advertiser’s payment. Therefore, each advertiser’s bid does not affect its own payment. This amounts to a second price auction (Vickrey 1961, Clarke 1971, Groves 1973). It is well known and not difficult to see that advertisers bid truthfully and advertiser inclusion is efficient.

4. A model of ad ordering with heterogeneous viewers

If television ads can be targeted to individual viewers, the algorithm in the previous section can be used to select, order, and price ads on the basis of viewer-specific continuation probabilities. This could be feasible in online video advertising, addressable advertising delivered over digital cable systems, or digital ad insertion in recorded programs stored on DVR hard drives. However, networks and cable systems currently deliver television ads en masse to many households simultaneously, and will continue to do so for the foreseeable future.
We now extend the previous model to assume that the television audience contains \( s = 1 \ldots S \) demographic segments. We use \( v_s \) to represent the proportion of audience in segment \( s \) at the start of the break, with \( \sum_{s=1}^{S} v_s = 1 \). We use \( c_a^s \) to represent the continuation probability of viewers in segment \( s \) after being exposed to ad \( a \). Thus, if ad \( a \) is placed in the first slot, we will have \( c_a^s v_s \) viewers in segment \( s \) remaining at the beginning of the second slot.

We assume advertisers differ in their valuations of reaching viewers in each segment. We assume that each advertiser may enter \( b_a^s \), its maximum willingness to pay to contact one viewer in segment \( s \), and will be charged proportionally for the number of viewers in that segment contacted.

We can now write the fraction of viewers in segment \( s \) watching at the start of slot \( j \) as

\[
P_j^s = \prod_{i=1}^{j-1} c_a^i.
\]  

(7)

Conditional on an ad sequence, network revenue is given by

\[
R = \sum_{j=1}^{J} \sum_{s=1}^{S} v_s P_j^s b_a^s.
\]  

(8)

The network’s problem is to select a set of ads and an ad ordering to maximize (8).

Ideally we could arrive at a simple pairwise ordering heuristic like inequality (3) in the case of homogeneous viewers presented in section 3. However, we can quickly see that this will be very difficult. Consider a simple example with \( S = 2 \) segments with \( v_1 = v_2 = 0.5 \), three slots, and three ads. Suppose \( b_a^s = 1 \) for all \( a \) and all \( s \). Each ad is represented by a vector of continuation probabilities \( \vec{c}_a = (c_a^1, c_a^2) \), with \( \vec{c}_1 = (1,0) \), \( \vec{c}_2 = (0,1) \), and \( \vec{c}_3 = (0.5,0) \). Given this set of continuation probabilities, the optimal ordering is clearly \( a_1, a_3, a_2 \). Once ad 1 has been shown, ad 2 will produce no revenue, so ad 3 should precede ad 2. However, if we instead had \( \vec{c}_3 = (0,0.5) \), then the optimal ordering would be \( a_2, a_3, a_1 \). Thus, the pairwise ordering between ads 1 and 2 depend on the continuation probabilities associated with ad 3.

We define a **heuristic** as a rule that, for every pair of ads \( a, a' \), determines which of \( a, a' \) comes before the other just by looking at \( a \) and \( a' \) without taking the other ads into account. We
define an input as a number of slots \( J \), a set of viewer segments with sizes \( \{ v_s \}_{s=1,...,S} \), a set of \( n \geq J \) ads with continuation parameters \( \{ c_s \}_{s=1,...,S} \), and a complete set of bids \( \{ b_a \}_{a=1,...,n} \).

**Proposition 1.** Assume that \( S = 2 \) and \( b_a^* = 1, \forall a, \forall s \). For every heuristic, there is an input on which that heuristic performs at least \( (k-1) \min (\frac{v_1}{v_2}, \frac{v_2}{v_1}) \) times worse than the best ordering.

**Proof.** We consider inputs that contain the following two ads: \( \bar{c}_1 = (1,0) \) and \( \bar{c}_2 = (0,1) \). Consider two mutually exclusive cases:

1) If the heuristic in question puts ad 1 prior to ad 2, then assume the input contains ads 3,4,..., \( J \) each with profile (0,1). The heuristic must also therefore sequence ad 1 prior to ads 3,4,..., \( J \) since this latter set is indistinguishable from ad 2. The total revenue produced by this ordering is equal to \( v_1 \). The optimal ordering would contain ads 2,..., \( J \) (in any sequence) followed by ad 1 and would produce total revenue of \( (J-1)v_2 \).

2) If the heuristic in question puts ad 2 prior to ad 1, we can apply a symmetric argument giving ads 3,4,..., \( J \) profile (1,0). Q.E.D.

Proposition 1 does not claim that every heuristic will fail in every situation. However, it shows that it is very easy to construct simple inputs on which any heuristic will fail terribly. If a heuristic can perform so poorly in a relatively simple case of 2 segments and reservation price homogeneity across advertisers and segments, it is likely to perform far worse in a more complex situation.

The heuristic fails because it only considers local properties of the input when making pairwise comparisons. Henceforth, we abandon the heuristic approach and seek to develop an algorithm that will consider properties of the input as a whole.

4.1. A simple algorithm
We propose here a simple algorithm to replace the heuristic notion considered above. This algorithm does not necessarily perform well, but will be useful to interpret our proposed algorithm below.

**Proposition 2.** Consider the following algorithm:

1) apply the pairwise comparison ordering rule (3) and dynamic program sorting rule in section 3 S separate times, considering only one segment in each application;
2) keep the ad selection and ordering that produced the highest revenue.

This algorithm will yield at least $S^{-1}$ of the optimal revenue possible, given any input.

**Proof.** To establish the minimum performance of this algorithm, we consider the input on which it does maximally poorly relative to the optimum revenue available. It is easy to show that the performance of the algorithm in proposition 2 increases with both advertiser preference heterogeneity, and with viewer segment size heterogeneity, so we assume homogeneity: $b_a^s = 1, \forall a, \forall s$ and $v_s = S^{-1}, \forall s$.

Consider an input that contains:

* set 1: at least $J$ ads which keep almost all viewers in all segments, that is, $c_{a}^{s'} = 1 - \varepsilon, \forall s, \varepsilon > 0$, and
* set 2: at least $J$ ads that are specifically targeted toward each segment $s' = 1, ..., S$:

$$c_{a}^{s'} = 1$$ and $$c_{a}^{s''} = 0, \forall s \neq s'.$$

The optimal selection and ordering would be to use the ads in set 1 and produce revenue of $J(1-\varepsilon)^{s'}$. The algorithm proposed in Proposition 2 will select the group of ads in set 2 targeting the largest segment $s'$ and will yield revenue of $JS^{-1}$. Therefore, as $\varepsilon$ approaches zero in the limit, the proposed algorithm will yield revenues $S^{-1}$ as large as the optimum. Q.E.D.

Note that the algorithm proposed in Proposition 2 does not fit into the class of heuristics because its output will be the optimal choice among sortings conditional on a given selection of ads. We mention it here only for expositional purposes, as it facilitates the explanation of the Audience Value Maximization Algorithm in section 4.2.
4.2. The Audience Value Maximization Algorithm

The easiest way to think about the AVMA is geometrically. Suppose we plot each advertisement \( a \) at a point \( \tilde{q}_a \) in \( R^s_+ \), where \( q_a^s = \frac{1-c_a^s}{b_a^s} \). Thus, ads will be closer to the origin when they retain more viewers and when they are associated with higher bids for the viewers they tend to retain. We illustrate this in Figure 2 for \( S = 2 \). The sortings produced by the algorithm in Proposition 2 will correspond exactly to the projections of all points on the coordinate axes, sorting from closest to the origin to furthest away.
We can generalize this idea while incorporating multiple segments at once by considering this projection onto other lines, rather than just the coordinate axes. Specifically, the AVMA has three steps:

1) For some line in the positive orthant, project all advertisements onto that line, and run the dynamic programming algorithm in section 3.2 to calculate the maximal revenue available on that line.

2) Repeat (1) for many lines in the positive orthant.

3) Select the line that produces the maximal revenue.

This algorithm has two subtleties to consider. First, as there are an infinite number of lines available in a plane, we need to discretize the advertisement space, for example in increments of one degree in each dimension, in order to check only a finite number of lines. For a given discretization, the number of lines will grow exponentially in the number of viewing segments, suggesting a tradeoff required between a larger number of viewing segments and a larger degree of approximation in the optimal solution.

Second, the dynamic programming algorithm needs to be modified, since we now must track audience sizes in each segment. We describe this in detail in the Appendix.

The pricing component is similar to before. A primitive of the model is a segment-specific bid from each advertiser. To determine each advertiser’s payment, we calculate
aggregate advertiser welfare conditional on optimal ad selection and ordering with advertiser \(a\) in slot \(j\), \(W(A)\). We then calculate aggregate advertiser welfare conditional on optimal ad selection and ordering under the exclusion of advertiser \(a\), \(W(A \setminus a_j)\). \(a\)’s total payment is the welfare its ad’s inclusion creates, \(p_a = W(A) - W(A \setminus a_j)\), conditional on this being less than its aggregate bid \(\sum_x b^*_a c^*_a\). Since \(a\)’s payment does not depend directly on its bid, the VCG nature of this auction ensures that the advertiser does not have a dominant strategy to deviate from entering a truthful aggregate bid. The network could modify this approach to return some small fraction \(\omega\) of the welfare created by including ad \(a\) to the advertiser, \(\sum_x b^*_a c^*_a - \omega[W(A) + W(A \setminus a)]\) in order to make truth telling a strictly dominant strategy.

Gross advertiser welfare is then given by

\[
R = \sum_{j=1}^{J} \sum_{s=1}^{S} v_j c^*_j b^*_j. \tag{9}
\]

While the AVMA is very intuitive, its complexity prohibits an analytical evaluation of its performance. We are currently using market data to estimate the primitives of our input to gauge how much the AVMA performs relative to networks’ current practices in the context of the typical commercial break.

5. Extensions, Challenges, and Limitations

The Audience Value Maximization Algorithm contains a static auction. It is thought that a failure to allow for advance selling or audience bundling explains Google’s recent decision to abandon its attempt to become a major broker of radio advertising time (Vascellaro 2009). In this section we discuss how to extend our proposed algorithm to allow for advance selling and audience bundling.

5.1. Advance Selling and Yield Management

Consider that advertisers may vary in their “time of arrival” to purchase advertising. For example, a movie studio may require an advertising blitz to promote the release of a new movie in 6 months’ time. Or, a politician responding to recent news reports may require last-minute advertising time to swing an election. This is an important practical issue in television markets.
The May/June “upfront” market allocates the majority of advertising time (60-80%) airing during the following September-August time period. The “scatter” market then allocates remaining ad time to advertisers up until the ads’ actual air dates.

It may well be the case that advertisers’ valuations are correlated with their time of arrival. It also seems likely that the number of potential advertisers seeking to buy ad time cannot be known with certainty by the television network. Since advertising stock is perishable, television advertising is a classic example of an industry in which yield management pricing practices may be profitably applied (Kimms and Muller-Bungart 2007). Indeed, several television networks (including NBC and MTV) already use yield management systems to price ad inventory.

If bids can be held until the complete set of bids is collected, the AVMA can be applied without modification. However, it would be unreasonable to require an advertiser to wait months to find out whether a particular advertising bid was accepted. We therefore assume that the game works as follows: within a time period, an advertiser arrives with a known continuation probability vector and enters its bid. The network must then either accept or reject the bid within the same period. We further assume that the network may choose to wait to reveal the advertiser's position and actual payment until the ad airs. We also assume that in the case of an unsuccessful bid, the advertiser is barred from entering additional bids in future periods in order to prevent bidders from "bidding up,” starting with zero and bidding incrementally until they discover the minimum possible bid that the network would accept.

We discuss here how to modify the static algorithm to accommodate yield management considerations. For simplicity, we take the definition of time period as an interval of time in which one potential advertiser "arrives" and enters a bid. Therefore we have uncertainty only over the number of bidders, not separate distributions over the number of bidders and the arrival rate.

Consider first the simpler case where the number of bidders is known with certainty. If we have one slot and two bidders, the first bid for the slot should be accepted if it is higher than the expected value of the second bid. The key question, then, is how to formulate this expected value; however, networks have ample historical data available to them to help guide this formulation. In the case of \( J > 1 \) slots, with \( n > J \), the trick is to accept each bid if and only if it is above the expected value of the \( J^{th} \) order statistic of the set of \( n \) bids.
This becomes a bit more complicated when we have uncertainty over the number of potential bidders $n$. The expected value of the $k$th order statistic now has to be integrated over a probability distribution for $n$.

The discussion above shows how to select which bids to accept. Since the ordering and pricing is not done until the complete set of bids has been collected, the rest of the algorithm may proceed as before.

5.2. Bundling

It is often the case that large television advertisers receive quantity discounts in exchange for purchasing bundles of audiences. These bundles may even span multiple networks within a corporation; e.g., NBC may bundle some of its broadcast audiences with those on its USA cable network. This practice could stem from large buyers exercising market power, reducing their transaction costs, or both. It seems worthwhile to consider how a network might implement bundling in the context of the AVMA.

Since the ad auction accommodates segment-specific advertiser bids, a large advertiser could deliver a single set of segment-specific bids to the network. The network could then compare these bids against the expected values of the order statistics defined in section 6.1 for its entire inventory of programs. This comparison would produce a set of programs which would best match the advertiser’s preferences. The network could consider both its aggregate profit across the entire inventory of programs when including the advertiser’s bids in this set of auctions, and its aggregate profit without those bids. Some fraction of this difference in aggregate profit could then serve as an upper bound for a fixed payment from the network to the advertiser to reward the advertiser for purchasing a large bundle of programs.

5.3. Promos and Branded Entertainment

It often is the case that networks must make allocate ads that are not priced by the market. For example, the network routinely decides not only how many ads for upcoming programs to include within a break, but also which upcoming programs to promote. And when making branded entertainment deals, networks have to decide whether to bundle commercials with product placements and what price to charge for the bundle. The expected values of bid order statistics discussed in section 6.1 could help the network optimize these decisions.

5.4. Implementation Challenges: How do we get there from here?
The AVMA is a practical algorithm that could be implemented relatively quickly. However it assumes the existence of several primitives which may not be immediately available. A network seeking to implement the AVMA would first need to set up four systems:

1) The AVMA assumes a complete set of distinct viewing segments. The definition and number of segments would have to be set at the network level and shared by all bidding advertisers.

2) The AVMA requires regular measurement of advertisement- and segment-specific continuation probabilities. Nielsen’s PeopleMeter panel is the natural place to start with this measurement as it is geographically randomly selected and will triple in size to about 100,000 people by 2011 (BroadcastEngineering 2007). However, participation requires a household’s consent and frequent participation (as each viewer must indicate her presence once or twice per hour), implying that it is nonrandom. Digital cable and DVR service have much larger sample sizes, but are not randomly assigned, implying a possible selection bias. Data fusion techniques could be used to combine the two sets of ratings to improve their accuracy.

3) The AVMA also requires a system to predict continuation probabilities with reasonable accuracy. Conditional on (2) for a large number of audiences and commercials, it seems likely this could be done using predictions based on historical information only. It might be desirable to supplement those predictions for new ads with other measures of viewers’ acceptance of advertising, such as ad feedback generated by websites like hulu.com. In addition, it could be possible to develop standardized testing procedures similar to many copytesting procedures already in use.

4) Advertisers must be educated about the benefits of the AVMA as its use requires some adjustments in their business processes. For example, they would receive regular feedback on their ads’ continuation probabilities. And advertisers would ideally use that feedback to edit their ads, in order to continually improve their ads’ continuation probabilities. Advertisers would further need to become comfortable with the budgeting and bidding implications of the new system.

5.5. Extensions

Future research might consider several interesting extensions. First, we have downplayed likely important factors such as how to predict advertiser- and segment-specific continuation probabilities. These are likely to vary with such factors as time of day, slot within the break, and repetitions of a given advertisement. Second, it would be fairly straightforward to modify the
algorithm considered here to allow for slot-dependent continuation probabilities and an endogenously chosen number of commercials to air within a break. A more challenging extension would be to decide on the optimal number of breaks within a program in addition to optimal numbers of commercials within each break. Doing this would require either extensive data or extensive assumptions to determine how viewership varies across the number of breaks within a program, and when the program would accommodate such breaks. A third direction for future research would be the problem discussed in section 5.3: given a menu of programs, how does the network optimally cross-promote its programs to maximize audience value at the network level? This could even be couched within a model of optimal program scheduling to automate the network scheduling decision. Finally, a very challenging direction for future research would be to consider competitive reactions in AVMA-related decisions.
Appendix. Formal Specification of the AVMA.

In this appendix we provide more formal specifications of the algorithms used to select and order ads in the AVMA. \( \gamma \) is a small chosen constant which defines the resolution of the different lines used for projection. We first describe the projection algorithm which is used to find the best ordering of selected ads.

**Algorithm 1**  
Projection Algorithm

1. for \( \alpha_s = 0, \ldots, \pi / 2 \) \( \mod \gamma \) in increments of \( \gamma \) do
2. for \( s = 1, \ldots, S - 1 \) do
3. if \( S = 2 \), let \( x_s = \cos(\alpha_s) \), else \( x_s = \cos(\alpha_s) \prod_{j<s} \sin(\alpha_j) \)
4. let \( x_s = \prod_{j=1}^{s-1} \sin(\alpha_j) \)
5. for each ad \( a = 1, \ldots, n \) do
6. let \( t_a = \sum_{s=1}^{S} (h_a^s)^{-1} x_s v_j (1-c_a^s) \)
7. sort the ads by increasing \( t_a \) 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 (9) 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.

Next, we describe the Dynamic Programming Algorithm used as a subroutine. We assume that the ads are sorted according to Algorithm 1 above. Algorithm 2 has a parameter \( \delta \) which impacts how accurately it performs. The closer to zero \( \delta \) is, the better the performance, but the longer the running time. A pre-processing step rounds down all probabilities \( c_a^s \) to the closest power of \( 1 - \delta \). Similarly, all values of \( v_j \) are rounded down to the closest power of \( 1 - \delta \). Thus, there are \( m = \log_{1-\delta} \frac{\delta^J}{J} \) different powers of \( 1 - \delta \) to which values could be rounded.

The key of the algorithm is to maintain a Dynamic Programming Table \( h \), which has \( S + 2 \) dimensions. It contains, for each tuple \( < a, j, q_1, \ldots, q_s > \) the maximum revenue that can be obtained using ads \( a, \ldots, n \) (without changing their order) in the slots \( j, \ldots, J \) under the assumption that \( q_s \) viewers of segment \( s \) are still watching at slot \( j \). We use the table \( r \) to record the inclusion of ads in optimum subsolutions for convenient output later on.
The important part is the Dynamic Programming loop where we compute the optimum solution for each subproblem by either including ad \( a \) in slot \( j \), or not including it. One important note: the best way to store the arrays \( r \) and \( h \) is probably to index them not with the actual \( q_s \) value, but rather with \( \log_{s^{-1}} q_s \), which takes on integer values \( 0,1,...,m \).

The running time of the algorithm is \( O(kn m^s) \). It thus grows exponentially in \( s \). Since \( m \) grows as \( \delta \) becomes smaller, \( \delta \) must be chosen to balance accuracy with computation time.

**Algorithm 2** Dynamic Programming Algorithm

1. {Pre-processing of probabilities}
   - for each ad \( a \) and each segment \( s \) do
     - if \( c_a^s < \delta / J \) then
       - round \( c_a^s \) down to 0
     - else
       - round \( c_a^s \) down to the nearest power of \( 1 - \delta \)
   - for each segment \( s = 1,...,S \) do
     - if \( v_s < \delta / k \) then
       - round \( v_s \) down to 0
     - else
       - round \( v_s \) down to the nearest power of \( 1 - \delta \)

2. {Initialization of DP array}
   - for all \( a, j, q_1,...,q_s \) do
     - let \( h[a,J+1,q_1,...,q_s] = 0 \)
     - let \( h[n+1,j,q_1,...,q_s] = 0 \)

3. {Central part of DP loops}
   - for \( j = J \) downto 1 do
     - for \( a = n \) downto 1 do
       - for each \( q_1,...,q_s = 1,1 - \delta,...,(1 - \delta)^{m(J+1-j)},0 \) do
         - if \( h[a+1,j,q_1,...,q_s] > h[a+1,j+1,q_1c_a^1,...,q_sc_a^s] + \sum_s b_a^sc_a^s q_s \) then
           - let \( h[a,j,q_1,...,q_s] := h[a+1,j,q_1,...,q_s] \)
           - let \( r[a,j,q_1,...,q_s] := 0 \)
         - else
           - let \( h[a+1,j,q_1,...,q_s] := h[a+1,j+1,q_1c_a^1,...,q_sc_a^s] + \sum_s b_a^sc_a^s q_s \)
           - let \( r[a,j,q_1,...,q_s] := 1 \)
   - {Post-processing to reconstruct the solution}
   - Let \( j = 1, a = 1, \) and \( q_s = v_s \) for each \( s = 1,...,S \)
while $j \leq J$ do
  if $r[a, j, q_1, \ldots, q_s] := 0$ then
    let $a := a + 1$
  else
    Output “include ad $a$ in position $j$”
  for each $s = 1, \ldots, S$ do
    let $q_s := q_s c_s^j$
  let $j := j + 1$ and $a := a + 1$
Works Cited


1 C3 is computed as the average program rating in all of the clock-minutes that contained any part of an ad and includes time-shifted DVR viewing up to three days after the programs’ airdate.

2 For comparison, consumers spent an average of 29:15 hours using the internet (up 4.6% over the previous year), including 3:00 hours watching internet video (up 53.2%).

3 There are a few exceptions in which prices depend on ad slot, such as the Super Bowl.

4 Other strategies are even more prevalent. Krugman, Cameron, and White (1995) observed viewers in their homes and found that subjects avoided watching the screen 67% of the time during commercial breaks. Tse and Lee (2001) called viewers immediately after commercial breaks and found that 81% reported engaging in some form of advertisement avoidance.

5 There is some evidence that fast-forwarded advertisement exposures are partially effective. Stout and Burda (1987) found that zipped ads produced brand recall effects about 20% as large as unzipped ad exposures. Gilmore and Secunda (1993) found that high-speed exposures to TV commercials boost prior learning, suggesting that zipped commercials can function effectively as reminder ads.

6 This is dynamic programming in the sense that a naïve approach to solving this problem would solve many overlapping subproblems. Constructing the dynamic programming table in this way requires us to solve each subproblem only once.

7 Alternatively, we might assume the network can estimate the advertiser’s continuation probability based on its ads’ historical performance, or assume the network requires the advertiser to commit to a set of minimum continuation probability thresholds.