INSEAD

The Business School for the World

Faculty & Research Working Paper

Stochastic Revenue Management Models for Media Broadcasting

Victor ARAMAN
Ioana POPESCU

2007/05/DS (revised version of 2005/43/DS)

Stochastic Revenue Management Models for Media Broadcasting

by

Victor Araman *

and

Ioana Popescu**

* Assistant Professor of Information, Operations and Management Sciences Leonard N. Stern School of Business, Kaufman Management Center, 44 West 4th Street, Room 8-74, New York, NY 10012

** The Booz Allen Hamilton Term Chaired Professor in Strategic Revenue Management, Associate Professor of Decision Sciences at INSEAD, Boulevard de Constance, 77305 Fontainebleau Cedex

Submitted to *Management Science*
manuscript MS-00544-2005

Stochastic Revenue Management Models for Media Broadcasting

Victor Araman
NYU Stern School of Business

Ioana Popescu
INSEAD

An important challenge faced by media broadcasting companies is how to allocate limited advertising space between upfront contracts and the scatter market in order to maximize profits and meet contractual commitments. We develop stylized optimization models of airtime inventory planning and allocation across multiple clients under audience uncertainty. In a short term profit maximizing setting, our results suggest that broadcasting companies should prioritize upfront clients according to marginal revenue per contracted audience unit, also known as CPM (cost per thousand viewers). For inventory planning purposes, accepted upfront market contracts can be aggregated across clients. The upfront market inventory should then be allocated to clients in proportion to contracted audience. Closed form solutions are obtained in a static setting. These results remain valid in a dynamic setting, when considering the opportunity to increase allocation by airing make-goods during the broadcasting season. We provide comparative statics that characterize the impact of contracting parameters, time and audience uncertainty on profits and inventory decisions. The results hold under general audience and scatter market profit models, as well as under service constrained models.

Key words: revenue management; stochastic inventory management; dynamic programming; advertising

1. Introduction

This paper is among the first to investigate, and provide formal models for profitably managing advertising space for media broadcasting. We begin by describing how advertising is sold, to illustrate the complexities of the media problem. This introduction further highlights our contributions, and related literature.

1.1. The Media Broadcasting Advertising Market

In the 1930s, the sponsorship of radio serials by makers of household-cleaning products led to the soap opera. Listeners were enthralled by episodic, melodramatic storylines, and advertisers were guaranteed a big audience. According to TNS media intelligence, U.S. advertising spending is expected to exceed USD 150 billion in 2006. Of this, television accounts for 44%, with advertisers ready to pay up to half a million dollars for a 30 second commercial in a popular show such as *Friends*, and new-show campaigns reaching up to USD 10 Million (source adage.com). According to the A.C. Nielsen Co., the average American watches more than 4 hours of TV each day, 30% of which is advertising, so you may wonder why viewers don't consider getting paid for it.

Media broadcasting companies such as TV channels, cable networks or radio stations collect most of their revenues from selling impressions, or "eye-balls", through advertisement space (30 second commercial slots) during various programmes. Typically, in north America and several European countries, the bulk of advertising space (about 80%) is sold during an *upfront market*, following the announcement of program schedules and prices for the year. The upfront market occurs during less than a couple of weeks in May in the US, much before the broadcasting season starts in mid-September. During this period, a few (about 20%) major advertisers buy long term contracts (media plans, or campaigns) at relatively low margins. These contracts are specified on a *CPM* (cost per thousand viewers) basis, and provide audience guarantees to the buyer. The remaining advertising space is sold throughout the broadcasting season on the *scatter (spot) market*, on a price-per-slot basis, usually at a higher margin and with no audience guarantee.

The upfront market is a buyers' market where clients hold a strong bargaining power. Upfront contracts stipulate a budget to be spent throughout the year, and a CPM, the ratio of the two providing a target level of audience (or ratings) to be reached by the campaign. Audience is unknown ex-ante, when inventory allocation is made, but provided ex-post by media rating agencies. During the upfront market, the broadcasting company (BC) provides major advertisers (MA) with an initial media plan (or sales plan) as part of the negotiated contract.[1] This allocation cannot be reduced if the plan over-delivers. In this case, the advertiser does not owe anything to the BC, who stands to lose revenue opportunities from the spot market. If the plan appears to under-deliver, then the BC can allocate additional slots for the MA, so called "make-goods" (or audience deficiency units, ADU), during the broadcasting season, to avoid under-performance penalties and maintain the client relationship.

The upfront and scatter markets are governed by different objectives. The main strategic objective for media companies is long term profitability, achieved through retention of key accounts in the upfront market. [2] On the short run, the objective of the media revenue management problem is optimal allocation of limited advertising space between major upfront contracts and the spot market throughout the year, to hedge against audience uncertainty, honor client contracts and maximize short term profits. Advertising capacity is fixed, by physical time limitation and regulations.[3]

[1] Sales are either direct or managed by intermediary buying and selling houses which represent their respective clients' interests. Our results and analysis are relevant to the party in charge of the inventory planning and revenue management function, for simplicity referred to as broadcasting company (BC).

[2] This partly explains the large volume sold at low margins upfront. Other arguments include stock market signaling, account executive incentives (sales volume), BC's risk aversion.

[3] In Europe for example, EU directives restrict advertising to 15% of total airtime, with at most 12 advertising minutes per hour, and not more than one break per movie. American regulations are less strict (about 30% of airtime).

The media problem is a highly complex multi-level problem. Industry practice, organizational and technical considerations, all argue for a hierarchical planning approach, separating strategic, operational and tactical layers of decision making. First, during the upfront market, the BC negotiates contract allocation and estimates overall inventory requirements for the upfront clients; this is known in practice as the strategic planning phase. Subsequently, account executives allocate an initial portion of advertising inventory to the client, with the provision that additional inventory, so-called make-goods, will be allocated during the scatter market if the plan under-performs. Sequential make-goods allocation decisions are made periodically during the rest of the year. This paper focuses on the inventory planning decisions upfront, and during the broadcasting season.

Other relevant elements of the media problem that are not addressed here include rate card pricing, contract negotiation and design, and scheduling. The latter uses planning recommendations to determine the exact location of an ad in a break, accounting for product conflict and other scheduling constraints.

1.2. Contribution and Structure

This paper focuses on the strategic and operational levels of the media planning problem, across upfront and scatter markets. We propose stylized models that take into account key elements of the media problem and provide insights for the various layers of decision making. From a modeling standpoint, an important contribution is that we specifically account for prior uncertainty about audience ratings. We also specifically model the trade-off between spot and upfront allocation to multiple clients. The contributions and structure of the paper are further detailed below.

The next section provides further background about the media problem and business model, allowing to set up the terminology, model ingredients and main assumptions. We discuss models of audience uncertainty and describe the upfront and scatter market revenue models.

Section 3 investigates optimal upfront contracting and planning for multiple clients. We model this problem as a multi-variate non-linear stochastic program, and obtain close-form solutions. We show that the total inventory dedicated for the upfront market can be determined by solving a single aggregate allocation problem, and individual client allocations should be in proportion to the contracted performance targets. In addition, these results extend to provide upfront planning recommendations across multiple periods (airdates or quarters) with varying audiences and performance targets. These results provide the BC with insights that significantly simplify the upfront contract negotiation and planning task. For modeling purposes, these results allow us to simplify the analysis by focusing on aggregate planning models.

Section 4 investigates the impact of audience uncertainty on the model, profits and decisions. In the media problem, the *value* of supply is a priori uncertain, measured by audience ratings. A transformation of audience (i.e. supply value) uncertainty into uncertainty in client demand, measured in inventory units (30 second ad slots), allows to map the media problem into a traditional newsvendor problem (see e.g. Porteus 2000). This transformation sheds light on the impact of audience uncertainty on optimal ad space allocation, in particular on the surprising result that (stochastically) larger, or less variable, audiences do not necessarily insure lower inventory allocations.

The results described so far focus on static models that provide upfront planning recommendations. Section 5 shows how inventory decisions are operationalized, upfront and during the broadcasting season, considering opportunities for future "make-goods" allocation. We propose dynamic models for make-goods vs. scatter market allocation, where audience, modeled as a stochastic process, is revealed periodically. We first show that the client contracting and aggregation results obtained for static upfront planning (Section 3) remain valid for make-goods provisioning. The multi-client make-goods allocation problem thus reduces to an aggregate, one-dimensional dynamic programming model, which is much simpler to analyze. Our structural results show that initial inventory commitment to the upfront market should be minimal (so called "gapping"), and the optimal make-goods allocation policy is a monotone threshold type policy. Such models allow the BC to dynamically monitor inventory and performance, and make optimal trade-offs between scatter market profitability and fulfilment of upfront market contracts. Finally, several intuitive heuristics that approximate the optimal dynamic policy are proposed and evaluated.

Section 6 reveals the robustness of our insights under alternative service constrained models, and discusses service differentiation and long term profit maximizing strategies. The last section concludes by outlining applications of our model in other settings, such as manufacturing and the management of non-profit organizations, as well as opportunities for further research in media revenue management.

1.3. Related Literature

The marketing literature has extensively investigated the impact of TV advertising on the consumer and sales (see e.g. Kanetkar et al. 1992 and Lodish et al. 1995), but largely ignored the issue of airtime inventory planning. Despite its richness and complexity, the media revenue management problem has received limited attention in the operations literature as well. Chapter 10.5 of Talluri and van Ryzin (2004) provides a brief account of the media revenue management problem (this book is the most complete reference on revenue management to date). To our knowledge, the only published works dedicated to the topic are Bollapragada et al. (2002) and Bollapragada and Garbiras (2004), referring to models successfully implemented at NBC. They provide deterministic solutions for sales plans generation, for one, respectively multiple clients. These models are deterministic versions of those in our Section 3.3, that account for scheduling constraints (e.g. show mix and product conflict), but ignore spot market opportunity costs and audience uncertainty.

A recent working paper by Bollapragada and Mallik[4] (2006) investigates how a risk averse BC should allocate rating points between upfront and scatter markets, when audience and scatter market revenues are uncertain and independent (mainly modeled as uniform). Target revenue and value (revenue) at risk objectives are optimized in a static one period model that aggregates demand from each market, similar to our aggregate model in Section 3. They compare the optimal solution resulting from their models with a risk neutral one, and provide comparative statics with respect to audience parameters. Their sole decision variable is the total number of rating points sold in the upfront market. There is no specification of how this translates into actual inventory allocation, and how this allocation is "operationalized" across clients and over time. Our work complements theirs by answering these questions in a risk neutral context.

There are several industries where capacity is sold partly in a forward (advance purchase) market and partly on a spot market, such as electricity markets, cargo shipping, manufacturing etc. A growing body of literature, reviewed by Kleindorfer and Wu (2003), investigates inventory management in such settings. Wu and Kleindorfer (2003) develop a framework that integrates spot market transactions with supply chain contracting. In the first stage, multiple sellers compete and bid on capacity. In the second stage, the buyer decides how much to exercise from the contracts and how much to procure from the spot market. In a multi-period model, Araman and Ozer (2005) study optimal inventory allocation between a long term sales channel and a spot market. In the revenue management literature, works on dual channels include Savin et al. (2005), studying an application in the rental industry, and de Vericourt and Lobo (2006) analyzing non-profit operations. Work in this area focuses mainly on production models under demand uncertainty and supply contracts.

One feature that distinguished the media problem from the above literature, and much of the operations literature, is the uncertain value of supply: audience is contracted for, but only realized after airtime inventory allocation is made. This aspect makes our problem similar in spirit to production planning models under random yield, for which Yano and Lee (1995) provide an excellent review. Our static and dynamic formulations are similar to the random yield (or reject allowance) model with deterministic demand and one, respectively multiple production runs, as summarized

[4] We thank the editor for bringing this work to our attention.

on pages 318, respectively 321, of Yano and Lee (1995). The latter dynamic model falls in the class of multiple lotsizing in production to order (MLPO) with random yield problems, surveyed by Grosfeld-Nir and Gerchak (2004). Such models aim to satisfy a fixed initial demand target through a pre-specified number of production runs to minimize inventory (holding and penalty) costs. In each production run an input quantity (lot size) is decided and a random output (function of the input and bounded below it) is produced. In our case, the input decision is the number of ads to air for a client, and output is the audience (or ratings) for these ads.[5] Our setup differs from these particular random yield models in several respects, including the interpretation of the objective function, the absence of holding costs, the general audience distribution (vs. specific discrete, e.g. binomial yield) and the non-linear scatter market profit (corresponding to production costs).

Most of the extensive random yield literature differs from our setting by focusing on periodic review and random demand models. For such models, simple myopic policies have also been noted to perform well, in particular those based on transformations to a newsvendor model (see Sepheri et al. 1986, Bollapragada and Morton 1999). These are different from our transformation in Section 4.1, in that our transformed random variable remains exogenous, i.e. independent of the decision variable. Service constrained random yield models are studied by Bassok et al (2002), who provide a linear heuristic. Gupta and Cooper (2005) investigate the stochastic monotonicity of the profit function in response to changes in the yield distribution. Their results are different from our sensitivity results in Section 4, due in particular to the absence of holding costs in our model.

While our insights focus exclusively on the media setting, several of our results, such as the unprecedented multiple-client analysis, contribute also to the random yield literature.

2. Media Terminology, Model Ingredients and Assumptions

This section provides further background for the media problem and business model. We describe the main ingredients of our models, including upfront and scatter market contracting terms, as well as audience and performance measures.

2.1. Audience Models, Performance Metrics and Forecasts

An important contribution of this paper consists in specifically considering audience uncertainty in the media planning task. Hence an important prerequisite for our model and results is to understand ex-post audience metrics and motivate ex-ante forecast models of audience variability. We begin by discussing existing metrics and possible levels of forecasted uncertainty.

[5] Our multiplicative performance model corresponds to the stochastically proportional random yield model.

Audience is the gross sum of all media exposures (the number of impressions, or "eyeballs" watching a given show), regardless of duplication. This is unknown ex-ante, but provided ex-post by media rating agencies, such as Nielsen Media Research in the US, or Médiamétrie in France. Another popular metric is *GRP* (gross rating point), also known as *rating*, the percentage of the target audience reached by an advertisement.[6]Because the ratio of audience to GRP is a constant (the number of TV viewers/households), the two terms are used interchangeably at no loss of generality. Other metrics include *reach* (the percentage of individuals within a targeted market that receive the marketing message at least once) and *frequency* (the number of times the target consumer is exposed to the marketing message during a campaign).

Example 1. Suppose that a campaign consists of x ads, and each viewer has an "individual probability" q of seeing each ad. The *frequency distribution* of spots seen by an individual is Binomial (x, q), and *average frequency* for the campaign is $\nu(x) = xq$. Given the market fraction κ of the network, or show, *reach* is given by the number of people who view at least one ad: $\rho(x) = \kappa(1 - (1 - q)^x)$. Both κ and q are part of any standard media metric report. The resulting GRP for one advertisement is $GRP = \kappa q$ and for the entire campaign $GRP(x) = \kappa q x$. $\square$

To obtain ratings forecasts, the audience (or equivalently GRP) that will view an ad is modeled ex-ante as a positive random variable ξ with mean μ, standard deviation σ, distribution F and density f. Its realization, as well as relevant metrics, are provided ex-post by media rating agencies. For example, a subjective probability q that each of M potential viewers watches the show suggests a binomial model $\xi \sim Bin(M, q)$, which for large market sizes M could be approximated by a normal distribution. All our results are distribution independent.

The performance of a media plan consisting of x ads targeted to an uncertain audience ξ is modeled as $\Psi(x, \xi) = x\xi$. This multiplicative performance model essentially assumes that a constant, but *a priori* unknown fraction of the population views a constant fraction of the ads. Here ξ captures the high level prior uncertainty about show popularity (or strength), arguably the main driver of audience uncertainty. (Once a show is on, audience is relatively more predictable: more or less the same people watch *Friends* every week.) The multiplicative performance model is simple to work with and fairly general. Our sensitivity results for upfront and dynamic make-goods allocation extend for more general performance measures of the form $\Psi(x, \xi)$, with Ψ an increasing function of both arguments and concave in x. The concavity assumption is justified by "repetition

[6] For example, during the week of November 20, 2006, the ABC show "Desperate Housewives" topped the US household ratings at 13.5, meaning that 13.5% of the estimated 110.2M TV households in the US watched the show that week.

wearout", i.e. the diminishing marginal benefit of repeatedly reaching the same individuals.[7]

In reality there is more variability to audience than just show popularity; for example audience is different from week to week, or for new rather than replay episodes of a show.[8] The next example discusses how the multiplicative model can be viewed as a first order approximation of a more complex model that captures audience variability across air-dates.

Example 2. The total number of (possibly duplicated) impressions from airing x commercials with audiences $\xi_1,\dots,\xi_x$ on a show is the convolution $\xi^{(x)}=\xi_1+\cdots+\xi_x$. If the number of ads x is large and audiences are i.i.d., the central limit theorem insures that $\xi^{(x)}\sim N(x\mu,x\sigma^2)$. However, audiences for a given show are usually highly correlated, especially within a narrow time-frame. Assuming they are exchangeable (i.e. their joint distribution is not affected by permutations of the individual random variables), jointly normally distributed with mean μ, variance σ^2 and correlation ρ, we obtain that $\xi^{(x)}\sim N(x\mu,(1-\rho+x\rho)x\sigma^2)$. Hence, denoting $Z\sim N(0,1)$, we have:

$$\Psi(x,\xi)=\xi^{(x)}=x\mu+\sigma\sqrt{x(1-\rho+x\rho)}\,Z=x\left(\mu+(\xi-\mu)\sqrt{\frac{1-\rho}{x}+\rho}\right),$$

which is increasing concave in x. Because x is large and ρ is high, ignoring second order terms, we obtain $\Psi(x,\xi)\simeq x\left(\mu+\sqrt{\rho}(\xi-\mu)\right)=x\zeta$, where $\zeta=\sqrt{\rho}\xi+(1-\sqrt{\rho})\mu$. Hence we can approximate Ψ by a multiplicative model corresponding to a modified intensity ζ with additional point-mass at the mean. In particular, if audiences are perfectly correlated $\rho=1$, we obtain $\xi^{(x)}\sim N(x\mu,x^2\sigma^2)$, which has the same distribution as the multiplicative model $x\xi$. $\square$

2.2. Revenue and Business Model, Terms and Decisions

The BC's objective considered in this paper is (short term) revenue maximization from selling limited advertising space (30 second commercials during breaks in programs, a.k.a. airtime inventory) to upfront and scatter markets. We next describe the business and revenue models governing these two markets. For further details regarding business processes see Bollapragada et al. (2002).

Upfront Market. Following the announcement of program schedules and prices in May, clients contact the BC with upfront market requests to purchase advertising space in bulk, for the entire season. A typical request consists of a budget B for the entire year, and a negotiated CPM (cost per thousand viewers) C.[9] This translates into a target performance $N=B/C$, measured by the total

[7] Models of the form $\Psi(x,\xi)=h(x)\cdot g(\xi)$ reduce to a multiplicative model, by a simple change of variable; for example *reach* can be modeled as $\Psi(x,\xi)=\rho(x)\xi=(1-(1-q)^x)\xi$, where ρ is increasing and concave in x.

[8] Such models are further investigated in Section 3.3.

[9] For example, in 2005, US TV advertising budget for P&G, the largest US advertiser, was \$2.5 Billion, whereas Toyota's was half a million. American prime time CPM averaged between \$20-\$35 for TV, and \$8-\$10 for cable.

audience of the campaign, in the client's desired demographic (e.g. men between 18 and 49 years old). Audience is unknown ex-ante, but provided ex-post by media rating agencies. Additionally, the MA may specify performance targets for specific periods (quarterly or weekly weighings) and programs (e.g. show mix, prime time targets).

In response to the sales request, the BC provides a sales plan or proposal, consisting of a list of x commercials to be aired, by show and airdate; the specific break location is decided later, close to broadcasting time.[10] Of course, this allocation cannot exceed the advertising capacity, denoted by Q (maximum number of 30 second slots available in a given period). An important quantity throughout the paper is the so-called *GRP allocation*, defined by $w = N/x$. This corresponds to the ratings (i.e. realized audience level $\hat{\xi}$) for which the performance target N is exactly met by allocating x units of advertising space. Good rating forecasts are crucial for sales plan generation, and airtime inventory management.

Because upfront contracts offer performance guarantees, the BC bears the risk of audience uncertainty. If the plan exceeds contracted performance, i.e. ratings exceed GRP allocation, the BC receives no additional payment. On the other hand, the BC incurs penalties if at the end of a season a plan has under-performed (e.g. if large ratings are committed upfront for a show that turns out to be a miss). Our models account for unmet performance targets N via a constant unit penalty b, which amounts to a total penalty cost of: $b \cdot (N - x\xi)^+$. Linear penalties are standard in the industry (see Bollapragada et al. 2002, Bollapragada and Mallik 2006), and common in goal programming. Under-performance penalties are also a natural interpretation of strategic service constraints, whereby the BC commits to satisfy client requests within a given probability (or expected fraction). Service constrained models are proved to be equivalent to penalty models in Section 6.

In summary, the direct expected profit from allocating x airtime inventory units (advertising slots) to an upfront client with budget B and target performance N, is modeled as $B - b\mathbb{E}[N - x\xi]^+$, where ξ is the (uncertain) audience for one advertisement. This model does not account for the opportunity cost of allocating inventory to the scatter market, discussed next.

Scatter Market. As opposed to upfront contracts which are audience/CPM based, scatter market pricing is per advertising slot, with no audience guarantee. Broadcasting companies typically have to commit scatter prices in advance.[11] On top of these baseline prices, the BC applies discounts based on buyer characteristics (bundle, volume, loyalty) and inventory status. The realized scatter

[10] For concrete examples of a plan request and proposal, see Figures 3 and 4 in Bollapragada et al. (2002).

[11] Such list prices or rate cards are typically public, see for example `http://www.ftv-publicite.fr`

market profit from y available slots with random audience ξ is denoted $\Pi(y, \xi)$. This model allows for scatter market price and demand to be implicitly audience dependent, e.g. $\Pi(y, \xi) = p(\xi)S(y, \xi)$, where S is the number of ad slots sold at price $p(\xi)$ given inventory level y. If x slots are allocated to the upfront market, the expected profit from the remaining $Q - x$ slots on scatter market is $\pi(x) = \mathbb{E}_\xi[\Pi(Q - x, \xi)]$.

All our structural results hold under a general scatter profit model $\pi(x)$ that is decreasing and concave in x, accounting for diminishing marginal returns to the scatter market. For simplicity of exposition, the first part of the paper focuses on linear models $\pi(x) = p(Q - x)$, corresponding to a constant scatter market price p per advertising slot. This allows for close form solutions and clearer insights at no loss of generality.

Decision Making Layers. The following hierarchical decision making approach is consistent with industry practice, and aligned with the structure of the paper. At the strategic planning phase, the BC must decide which client contracts to accept upfront. For this purpose, the firm needs to assess how much inventory is required to satisfy upfront market clients. This inventory is not committed to the client upfront; it serves only for strategic planning purposes at the upfront stage. At the operational level, the decisions of how much initial inventory to commit to the upfront clients/market are made, with the specific provision of subsequent make-goods allocation during the scatter market. During the broadcast season, the BC periodically decides how many additional make-goods to allocate to upfront clients (vs. the scatter market), given the current performance of a campaign. The current practice in the industry is to make such decisions qualitatively, rather than using any formal models (see e.g. Bollapragada and Mallik 2006). We begin by investigating the strategic level decisions in the next section.

3. Contracting and Planning for Multiple Clients

During the upfront market, multiple clients approach the BC around the same time (during a couple of weeks in May in the U.S.) and negotiate advertising plans for the entire season. Requests consist of a budget B_i and CPM (cost per thousand viewers) C_i, resulting in a target performance $N_i = B_i/C_i$.

During the brief upfront market period, contracts are negotiated in parallel by account executives. These report to a strategic planning group, who oversees the process, and considers all client requests and proposals in order to provide strategic sales guidelines. This process motivates a joint optimization model (as opposed to a dynamic, sequential approach) for simultaneous contracting and inventory planning for multiple clients.

We begin by studying the optimal inventory planning problem for a set of contracted clients, and then fold back the results to answer which client contracts the BC should accept in the first place.

3.1. Multi-Client and Aggregate Planning

Assume that the BC signed contracts stipulating performance targets $N_i, i \in K$ with a set of k clients. In order to estimate the amount of inventory $X_i, i \in K$ necessary to satisfy each client's request at minimal cost to the firm, the following multi-client inventory planning problem is solved:

$$(M) \qquad c^*_M(\mathbf{N}) = \min_{X_i \geq 0, \sum X_i \leq Q} p \sum_{i \in K} X_i + b \sum_{i \in K} \mathbb{E}[N_i - X_i \xi]^+, \tag{1}$$

where $\mathbf{N} = (N_1, \ldots, N_k)$. Throughout the paper, bold characters are reserved for vectors.

The main trade-off captured by this model is between the opportunity cost of allocating a slot to the spot market and the penalty cost of not meeting the performance targets of upfront market clients. We assume a uniform under-performance penalty b across clients, consistent with a unique, strategic service level adopted by the company (see Section 6).[12]

Let $N = \sum_i N_i$ be the aggregate performance contracted upfront, and $X = \sum_i X_i \leq Q$, the total inventory provisioned for the upfront market for planning purposes. The main result of this section is that clients can be aggregated for planning purposes. Specifically, the total inventory allocation X to the upfront market equals the optimal allocation to a single "mega client" with aggregate target performance N, under identical scatter market conditions. This inventory is optimally divided among clients in proportion to their contracted performance targets N_i.

Aggregate Planning. Central to our derivations is an aggregate planning model, which cumulates performance targets across all accepted clients into the single cumulative target $N = \sum_i N_i$:

$$(A) \qquad c^*(N) = \min_{0 \leq x \leq Q} px + b\mathbb{E}[N - x\xi]^+. \tag{2}$$

Recall that F is the distribution of the audience ξ, and denote its left tail expectation by:

$$G(u) = \mathbb{E}[\xi I_{\xi \leq u}] = \mathbb{E}[\xi; \xi \leq u]. \tag{3}$$

This allows to write the cost objective in (2), i.e. the cost of allocating x units of inventory to the upfront, as:

$$c(N, x) = px + b\Big[NF(N/x) - xG(N/x)\Big]. \tag{4}$$

The next result follows from the first order conditions, by convexity of the cost function and monotonicity of G. Throughout the paper, increasing/decreasing refer to weak monotonicity.

[12] Differentiating penalties is technically feasible, but would not be appropriate in a short term profit optimization model such as ours (e.g. because it leads to prioritizing clients in reverse order of their importance).

LEMMA 1. *The optimal solution to Problem (A) is $x^* = \min(Q, \bar{x})$, where the unconstrained optimum $\bar{x}$ satisfies:*

$$G(N/x) = p/b. \tag{5}$$

Furthermore, x^* is piecewise linear and increasing in the target performance N, increasing in the penalty b and decreasing in the average spot price p, all else equal. The optimal cost $c^*(N)$ equals $bNF(G^{-1}(p/b))$, if $p/b \geq G(N/Q)$ and $c(N,Q)$ otherwise.

Multi-Client Planning. The next result shows that the aggregate model (A) actually produces the optimal recommendation for total upfront market allocation. This is surprising, because the aggregate performance target induces additional pooling effects that may not occur when penalties are incurred at the client level.

PROPOSITION 1. *The optimal solution to Problem (M) equates GRP allocation across clients:*

$$\frac{N_i}{X_i^*} = \frac{N}{X^*}, i \in K, \text{ where } X^* = \sum_{i \in K} X_i^* = \min(Q, N/G^{-1}(p/b)) \text{ and } N = \sum_{i \in K} N_i.$$

That is, the total upfront allocation X^* solves the aggregate Problem (A) with target performance N. Moreover, the optimal costs of the two problems are the same $c_M^*(\mathbf{N}) = c^*(N)$.

Proof: For any feasible multi-client allocation $\mathbf{X} = (X_1, \ldots, X_k)$, the corresponding aggregate allocation $X = \sum_{i \in K} X_i$ is feasible for model (A), and the corresponding cost can only be lower in the aggregate model (because of pooling effects $\sum a_i^+ \geq (\sum a_i)^+$). Hence,

$$c_M^*(\mathbf{N}) = \min_{X_i \geq 0, \sum X_i \leq Q} p \sum_{i \in K} X_i + b \sum_{i \in K} \mathbb{E}[N_i - X_i \xi]^+ \geq \min_{0 \leq x \leq Q} px + b\mathbb{E}[N - x\xi]^+ = c^*(N). \tag{6}$$

Consider now the optimal solution x^* of the aggregate Problem (A), and define $X_i^* = N_i x^*/N, i \in K$ This is feasible to Problem (M), and satisfies $\sum_i X_i^* = x^*$ and $\sum_{i \in K} \mathbb{E}[N_i - X_i^* \xi]^+ = \mathbb{E}[N - x^* \xi]^+$. So $\mathbf{x}^*$ achieves the same cost as the optimal aggregate model cost $c^*(N)$, hence it is optimal for model (M) (and unique by convexity of the objective). Moreover, the optimal costs of the multi-client and aggregate problems are equal. □

These results remain true for a general scatter profit $\pi(x)$, equating allocation revenues (as opposed to costs) for the multi-client and aggregate models.

In summary, given a set of clients and their respective requirements, the firm only needs to determine the upfront market GRP allocation, i.e. contracted audience per inventory unit $w = N/X^*$. This should be set equally across clients i.e. $N_i/X_i^* = w$. Bollapragada and Mallik (2006) use GRP allocation as a single decision variable in an aggregate upfront market model; our results in this section validate their approach.

3.2. Multi-Client Contracting and Priority Heuristic

We now step back to investigate which clients the BC should accept, among a given set of client requests, consisting of a target performance N_i and budget B_i, at the negotiated CPM rate $C_i = B_i/N_i$. We provide a simple heuristic for upfront market contract negotiation and client prioritization, which is consistent with industry practice. Our model is limited to the firm's short term profit maximization problem, and thus ignores the impact of current decisions on customer retention in the context of the firm's long term profit maximization problem (see Section 6.2 for a discussion).

We consider a setup where the BC may offer partial fulfilment of client requests, at the negotiated CPM level C_i. Let $y_i \in Y = [0,1]$ be the satisfied fraction of client i's demand,[13] and x_i the amount of inventory provisioned for client i. Accepting client i increases revenues by $B_i y_i$, less penalties for unmet performance, leading to the following revenue maximization model:

$$P = \max_{y_i \in Y} \max_{\sum x_i \le Q, x_i \ge 0} p(Q - \sum_{i \in K} x_i) + \sum_{i \in K} \left(B_i y_i - b\mathbb{E}[N_i y_i - x_i \xi]^+ \right). \tag{7}$$

Combining terms, this can be equivalently written as

$$P = \max_{y_i \in Y} p\, Q + \mathbf{B}'\mathbf{y} - c^*_M(\mathbf{N} \circ \mathbf{y}), \tag{8}$$

where $\mathbf{N} \circ \mathbf{y} = (N_1 y_1, \ldots, N_k y_k)$ denotes the componentwise product vector of $\mathbf{N}$ and $\mathbf{y}$. The last term stems from the multi-client allocation model (M) with contracted client performance targets $N_i y_i$. Proposition 1 reduces this to an aggregate model (A) with cumulative target $\mathbf{N}'\mathbf{y}$:

$$c^*_M(\mathbf{N} \circ \mathbf{y}) = c^*(\mathbf{N}'\mathbf{y}) = \min_{0 \le X \le Q} pX + b\mathbb{E}\left[\mathbf{N}'\mathbf{y} - X\xi\right]^+, \tag{9}$$

and the resulting optimal inventory X^* is optimally divided among clients in proportion to their contracted targets $N_i y_i$. Problem (7) can be restated as:

$$\begin{aligned} P &= pQ + \max\nolimits_{y_i \in Y} \mathbf{B}'\mathbf{y} - c^*(\mathbf{N}'\mathbf{y}) \\ &= pQ + \max\nolimits_{y} \left\{ -c^*(y) + \max_{\mathbf{N}'\mathbf{y} = y; y_i \in Y} \mathbf{B}'\mathbf{y} \right\}. \end{aligned} \tag{10}$$

For any given fixed $y = \mathbf{N}'\mathbf{y}$, the inner problem is a knapsack model, whose optimal fractional solution amounts to serving clients in decreasing order of their marginal profitability, or CPM, $C_i = B_i/N_i$.[14] Hence this will also be true for the optimal solution of Problem (10). Client demands should be served as long as they are profitable and inventory is available. In absence of capacity constraints, a client is profitable whenever CPM C_i exceeds expected marginal penalty $bF(G^{-1}(p/b))$. Hence, also in the capacitated case, no client should be accepted if $C_i < bF(G^{-1}(p/b))$.

Our results in this section can be summarized as follows:

[13] Restricting $Y = \{0,1\}$ corresponds to an all-or-nothing contracting model.

[14] Under all-or-nothing contracting, this is known as the greedy heuristic for the corresponding $0-1$ knapsack model.

PROPOSITION 2. *The following sequential procedure for multi-client contracting and planning is optimal under partial order fulfilment:*

1. *Accept client contracts in decreasing CPM order, as long as they are profitable, i.e.* $C_i \geq bF(G^{-1}(p/b))$*, and inventory is available.*

2. *Given the fractions of contracted client requests* $\mathbf{y}$*, solve aggregate model* (A) *with cumulative performance target* $\mathbf{N}'\mathbf{y}$ *to obtain total upfront market allocation* $X^* = \min(Q, \mathbf{N}'\mathbf{y}^*/G^{-1}(p/b))$.

3. *Allocate the upfront market inventory* X^* *among clients in proportion to contracted performance targets* y_iN_i*, i.e. by equating average GRP allocation* $N_i/X_i^* = N/X^*$.

The proposed sequential approach is practically appealing due to its simplicity and transparency, relative to the overall complexity of the media problem. Moreover, the insights are general and robust, and remain valid under general scatter market profit models $\pi(x)$. Section 5 shows that the key insights are also preserved when modeling the additional recourse provided by the opportunity of allocating additional inventory (make-goods) over time. These results allow to focus the analysis on the aggregate model (A).

3.3. Extension. Multi-stage Upfront Planning

The results of Proposition 2 further extend to account for upfront planning over multiple periods, with varying audience, as well as periodic (e.g. quarterly) performance targets N_i, also known as weekly/quarterly weighings.[15] Specifically, the inventory allocation problem with multiple (periodic) targets can be formulated as:

$$(MT) \quad \min_{\mathbf{x}\in[\mathbf{0},\mathbf{Q}]} \mathbf{p}'\mathbf{x} + b\sum_{i\in K}\mathbb{E}[N_i - x_i\xi_i]^+. \tag{11}$$

If clients demand a cumulative target N for multiple airdates with varying audiences, we obtain:

$$(CT) \quad \min_{\mathbf{x}\in[\mathbf{0},\mathbf{Q}]} \mathbf{p}'\mathbf{x} + b\mathbb{E}[N - \mathbf{x}'\boldsymbol{\xi}]^+ \tag{12}$$

PROPOSITION 3. *Assume that audience distributions* ξ_i *are exchangeable, scatter prices are constant* $p_i \equiv p$*, and ad space capacity is uniform* $Q_i \equiv Q$ *across periods.*

(a) The multi-target problem (MT) *reduces to solving an aggregate model* (A) *with cumulative target* $N = \sum_i N_i$*, and allocating this inventory among periods in proportion to the target* N_i.

(b) The cumulative target problem (CT) *reduces to solving an aggregate model* (A) *with audience* $\xi = (\sum_{i=1}^k \xi_i)/k$*, and allocating this inventory equally among periods.*

[15] See e.g. Table 10.7 in Talluri van Ryzin 2004 for an example of quarterly weighings.

The above result shows that, if audience is homogeneous (but not necessarily independent) across air-dates, and consistently priced on scatter (equal scatter market CPM p/μ), then inventory should be allocated so as to balance GRP allocation across periods. Incidentally, uniform allocation is standard industry practice, supported by clients' preference against burstiness. The result is technically appealing, as it reduces to solving one aggregate model of type (A), allowing to embed multi-stage planning as a last stage in the sequential aggregation procedure described in Proposition 2.[16] These results also extend for general scatter market profit models.

4. The Impact of Audience Uncertainty

One of the main contributions of this paper is to incorporate *stochastic* models of audience in the media planning problem. In that sense, it is relevant to characterize the impact of audience uncertainty on profits and managerial decisions. Our analysis provides decision support regarding what channels are better suited for various contracts, how to adjust client allocations in response to changes in channel parameters (such as audience variability and show popularity), and what are the costs of ignoring these changes (e.g. direction of change in revenues). Some of our results challenge conventional wisdom and intuitive practices. Specifically, we show that (stochastically) larger audience, or lower audience variability, generate higher profits but do not necessarily lead to lower allocations.

The previous section showed that the multi-client allocation problem reduces in general to solving the aggregate allocation model (A). Hence it is sufficient to focus on model (A) in order to analyze how audience uncertainty impacts the system and, more specifically, the optimal inventory allocation. Our approach relies on converting client demand from deterministic audience units (eyeballs) to stochastic inventory units (advertising slots). This allows to adapt traditional inventory theory insights and results to the media problem.

4.1. Supply vs. Demand Uncertainty. An Inventory Formulation

In the media problem, demand, measured in audience units, is deterministic (N), whereas the value of allocated supply is uncertain $(x\xi)$; this makes the problem akin to a random yield model. An alternative perspective on the problem emerges by translating all model ingredients, in particular demand, in terms of inventory units. This is particularly appealing because the BC's operational decision is necessarily in terms of ad space inventory.

[16] One can obtain similar results in terms of budget variables $v_i = p_i x_i$, if scatter market CPM distributions are homogeneous (ξ/p_i exchangeable) across air-dates, but audiences and prices may not be. This is particularly relevant when dealing with show mix constraints. We obtain that the firm should balance GRP allocation, prorated by scatter market price.

Because the left tail audience expectation $G(u)$ is increasing and bounded in $[0,\mu]$, the function $H(u) = G(u)/\mu$ can be viewed as the distribution function of a random variable ζ with density $h(u) = u/\mu f(u)$. This distribution captures left tail variability in the audience ξ. The likelihood ratio of the two measures is ξ/μ. Consider the random variable $\theta = N/\zeta$, with cdf $\mathbb{P}(\theta \le x) = \mathbb{P}(\zeta \ge N/x) = 1 - H(N/x) = \mathbb{E}[\xi; \xi \ge N/x]/\mu$. The cost function in Problem (A) can be written as:

$$c(N,x) = px + b\mu\mathbb{E}[\xi/\mu(N/\xi - x)^+] = px + b\mu\mathbb{E}_\zeta[N/\zeta - x]^+ = px + b\mu\mathbb{E}_\theta[\theta - x]^+. \tag{13}$$

By converting audience uncertainty ξ into demand uncertainty θ, measured in inventory units, the media problem is reduced to a traditional inventory model (see e.g. Porteus 2000) given by equation (13), with $cost = p$ and $price = b\mu$. The corresponding critical fractile solution, equivalent to (5), is given by

$$\mathbb{P}(\theta \ge x) = \frac{p}{\mu b}, \tag{14}$$

or in terms of GRP allocation $w = N/x$,

$$\mathbb{P}(\zeta \le w) = \frac{p}{\mu b}. \tag{15}$$

Inventory considerations suggest writing $x^* = x_0 + SS_\theta$, where SS_θ is the safety stock, and $x_0 = \mathbb{E}\theta = \mathbb{E}\big[N/\xi \cdot \xi/\mu\big] = N/\mu$ is the optimal allocation corresponding to a deterministic audience $\xi \equiv \mu$. Practical considerations suggest that penalties much exceed scatter market CPM, $b \gg p/\mu$, so the right hand side in (14) is positive, and close to 1. This indicates a positive safety stock.[17]

We conclude that a deterministic model prediction would typically underestimate the upfront allocation (i.e. overestimate GRP allocation) required to hedge for audience uncertainty, i.e. usually $x^* \ge x_0 = N/\mu$.

Example 3. For a binomial audience distribution $\xi \sim Bin(M,q)$, we have $\mu = Mq$ and

$$G(u) = \sum_{k=0}^{u} k\binom{M}{k} q^k(1-q)^{M-k} = Mq\sum_{k=1}^{u}\binom{M-1}{k-1}q^{k-1}(1-q)^{M-k} = \mu\sum_{k=0}^{u-1}\binom{M-1}{k}q^k(1-q)^{M-1-k}.$$

Because the cdf of ζ is defined as $H(u) = G(u)/\mu$, we obtain that $\zeta \sim 1 + Bin(M-1,q)$. Remark that ζ has (slightly) higher mean and lower variance than ξ, and both distributions stochastically increase with M and q. In practice, the market size M is very large, so ζ can be approximated by a normal distribution with mean $1 + (M-1)q$ and variance $(M-1)q(1-q)$. The first order condition (15) amounts to inverting the beta function that computes the binomial cdf, or the normal cdf for the approximation model. The latter allows to approximate the optimal GRP allocation as follows

$$w^* = N/x^* \approx \mathbb{E}\zeta + \sigma_\zeta\alpha = 1 + (M-1)q + \alpha\sqrt{(M-1)q(1-q)}, \tag{16}$$

[17] Formally, the safety stock is positive ($x^* > x_0$) if and only if $b > p/G(\mu)$, i.e. $H(\mu) > p/(\mu b) \approx 0$.

where α is such that $\mathbb{P}(Z \le \alpha) = \dfrac{p}{\mu b} = \dfrac{p}{Mqb}$ and Z is a standard normal distribution. In particular, $p \ll \mu b$ implies α is negative and sufficiently small, so $w^* < Mq$, as long as market size M is sufficiently large relative to viewing probability q. Hence a positive safety stock is required, $x^* = N/w^* > N/(Mq) = N/\mu = x_0$. □

4.2. Sensitivity to Changes in Audience Distribution

In order to measure the impact of audience uncertainty on profits and decisions, we set up an abstract factor model, where audience and spot market profit are driven by a common "show popularity" factor z. This can be seen as a signal or available information on the current status of the audience and market. The end of this section briefly discusses a model where z controls an increase in audience variability.

Consider the aggregate planning profit maximization model:

$$r^*(z) = \max_{0 \le x \le Q} r(x,z), \quad r(x,z) = (Q-x)p(z) - b\mathbb{E}[N - x\xi(z)]^+. \tag{17}$$

Assume that scatter market price $p(z)$ is increasing in z and audience $\xi(z)$ is stochastically increasing in z with respect to first order dominance. It follows that $r(x,z)$ and $r^*(z)$ increase with show popularity z. On the other hand, popularity does not necessarily decrease upfront allocation cost $c(x,z) = p(z)x + b\mathbb{E}[N - x\xi(z)]^+$. This is because spot market opportunity costs increase whereas penalty costs decrease.

It may appear intuitive that optimal allocation $x^*(z)$ should decrease as audience increases (in the sense of first order dominance), and we shall provide conditions for this to be true. The following counterexample shows, however, that this is not always the case. Somewhat surprisingly, what can be perceived as a higher popularity show may actually require higher allocation, and this cannot be attributed to inconsistent pricing. By consistent pricing we mean that scatter prices reflect popularity effects, by keeping scatter market CPM $p(z)/\mu(z)$ constant.

Counterexample 1. For a given positive constant A, consider two positive audience distributions ξ_1, ξ_2 with continuous cdfs F_1, respectively F_2, satisfying $F_2(x) = F_1(x)$ for all $x \ge A$ and $F_2(x) < F_1(x)$ for $x < A$. By construction, $\xi_1 <_{st} \xi_2$, so in particular $\mu_1 < \mu_2$. For all $x < N/A$, we have:

$$H_1(x) = \mathbb{P}(\theta_1 \le x) = \mathbb{E}[\xi_1; \xi_1 > N/x]/\mu_1 = \mathbb{E}[\xi_2; \xi_2 > N/x]/\mu_1 > \mathbb{E}[\xi_2; \xi_2 > N/x]/\mu_2 = \mathbb{P}(\theta_2 \le x) = H_2(x),$$

so in particular θ_1 is *not* stochastically larger than θ_2. The critical fractile condition (14) shows that, for equal scatter CPM $p_1/\mu_1 = p_2/\mu_2 < b\bar{H}_i(N/A)$, the resulting allocations satisfy $x_1^* < x_2^*$.

For example, let ξ_1 be uniform on the support $[1,3]$, so $\mu_1 = 2$. Define ξ_2 such that, with probability $\alpha = 0.5$, ξ_2 is uniform on $[1.5, 2]$, otherwise, it is uniform on $[2,3]$ (e.g. the show is a miss with probability α). These distributions satisfy the above conditions, so $\xi_1 <_{st} \xi_2$. However, for the same CPM=5 and penalty level $b = 10$, Eq. (14) gives capacity allocations $x_1^* = 22 < x_2^* = 23$. □

The critical fractile solution (14) indicates that optimal inventory allocation decreases in the popularity factor z, under a stochastically decreasing demand uncertainty, $\theta(z)$ (a sufficient condition). The problem is that a stochastically increasing audience distribution $\xi(z)$ does not guarantee inventory unit demand $\theta(z)$ to be a stochastically decreasing, as the above counterexample indicates. The result is true, however, for a variety of parametric classes of audience distributions, such as uniforms, exponentials, gammas and lognormals, with the natural order of parameters that induces first order dominance (see e.g. Table 1.1 in Müller and Stoyan 2002). In addition, the next result presents two general classes of audience distributions, which also insure natural comparative statics with respect to popularity effects.

Proposition 4. *The following alternative conditions insure that the optimal allocation $x^*(z)$ is decreasing in the popularity factor z. Let ξ_0 be a positive random variable with fixed distribution (independent of z), and $\mu(z)$ is an increasing function of z.*

(a) Audience is multiplicative $\xi(z) = \mu(z)\,\xi_0$, and scatter market CPM, $p(z)/\mu(z)$ is constant (or does not decrease with z).

(b) Audience is additive $\xi(z) = \mu(z) + \xi_0$, where ξ_0 is unimodal with mean 0 and non-negative mode. Moreover, penalties are sufficiently large to insure a positive safety stock, i.e. $b > p(z)/G(\mu(z))$.

Audience Variability. Consider now a model where the factor z in (17) is a measure of audience variability, instead of show popularity. Similar arguments show that revenues decrease with variability in audience, i.e. a show with more variable audience, as measured by the concave or increasing concave order,[18] yields stochastically lower profits. Revenue monotonicity holds for these orders, provided that $p(z)$ is also concave in the variability factor z. Again, it may appear surprising that higher audience variability does not necessarily require larger allocation.

Counterexample 2. Consider Example 3, where $\xi(z) \sim Bin(M(z), q(z))$, with decreasing viewing probability $q(z)$ but a constant expected audience $M(z)q(z) \equiv \mu$, so z increases the variance of the audience distribution ξ. Eq. (16) can be written as $w^*(z) \approx 1 + \mu - q(z) + \alpha\sqrt{(\mu - q(z))(1 - q(z))}$. Assuming that $p(z) \equiv p$ (constant scatter market CPM), Example 3 shows that indeed $\alpha(z) \equiv \alpha$. Hence GRP allocation $w^*(z)$ is increasing in z, i.e. the optimal allocation $x^*(z)$ decreases with higher audience variance. □

[18] These are standard variability orders, defined as follows: X dominates Y in the (increasing) concave order if $\mathbb{E}[u(X)] \geq \mathbb{E}[u(Y)]$ for all u (increasing) concave such that the expectation exists (see Müller and Stoyan 2002 p.16).

In general, we cannot say anything about monotonicity of the optimal allocation with respect to the standard deviation of the audience distribution. This is essentially because variance is not an appropriate measure of down-side risk, which is the relevant audience risk in the media problem.

5. Operational Decisions. Dynamic Make-Goods Allocation

Our results so far focused on static models that provide BCs with decision support during the upfront market decision process, including which upfront client contracts to accept and how much inventory to provision for each of these clients. These inventory provisions are for strategic planning purposes; they are not actually committed to the client upfront. This section shows how inventory allocation is operationalized. First, the BC decides how much inventory should be committed to the client upfront, with the provision that this allocation cannot be reduced, but can be increased by subsequently allocating make goods. Second, as the season starts, and audience ratings unfold, the BC further decides when and whether to increase initial allocation to under-performing upfront market client campaigns, by airing make-goods. This decision is continuously traded-off against immediate scatter market profit opportunities.

The goal of this section is to provide the BC with a decision support tool to dynamically monitor inventory and profits, and adjust allocation to upfront clients in the most profitable way. At any point in time, given the dedicated inventory and achieved performance of the client campaigns, our results diagnose whether a campaign is potentially under-performing. In this case, our model suggests how many make-goods should be aired to each client, and provides projections of future expected profits.

This section works with a general scatter market profit model $\pi(x) = \mathbb{E}_\xi[\Pi(Q - x, \xi)]$, as described in Section 2.2. For simplicity of exposition, we assume audiences are i.i.d over time $\xi_t \equiv \xi$ (our results extend for Markovian processes). Consistent with the broadcasting business cycle, the planning horizon T is assumed to be one year (discretized into weeks, months or quarters).

Let $\mathbf{x}_0$ denote the initial "irreversible" allocation committed to clients upfront. In each period $t \geq 1$, given the remaining client performance targets $\mathbf{N}_t$, the BC decides how many additional make-goods $\mathbf{x}_t - \mathbf{x}_0$ to allocate to upfront advertisers in order to maximize scatter market profits (realized each period) net of penalties for unmet performance, calculated at the end of the horizon. The resulting profit (net of contracted upfront client budgets $\mathbf{B}$) is denoted $V_t^M(\mathbf{x}_t, \mathbf{N}_t)$, and its optimal value $J_t^M(\mathbf{N}_t)$. For a contracted target $\mathbf{N}$, this leads to a dynamic programming model, that optimizes $J_0^M(\mathbf{N}) = \max_{\mathbf{x}_0} J_0^M(\mathbf{x}_0, \mathbf{N})$, given recursively by the following Bellman Equation:

$$
\begin{aligned}
J_t^M(\mathbf{N}_t) = & \max_{\mathbf{x}_0 \le \mathbf{x}_t; \mathbf{1}'\mathbf{x}_t \le Q} && V_t^M(\mathbf{x}_t, \mathbf{N}_t) \\
\text{where} \quad & V_t^M(\mathbf{x}_t, \mathbf{N}_t) = && \pi(\mathbf{1}'\mathbf{x}_t) + \mathbb{E} J_{t+1}^M(\mathbf{N}_t - \xi \mathbf{x}_t) \\
\text{and} \quad & J_T^M(\mathbf{N}_T) = && -b\mathbf{1}'\mathbf{N}_T^+,
\end{aligned}
\tag{18}
$$

where $\mathbf{1}$ denotes the vector of ones.

Problem (18) can be further embedded in the multi-client contracting problem studied in Section 3.2. The following model considers the full allocation recourse available to the decision maker in order to assess which contracts to accept initially:

$$
\max_{y_i \in Y} \sum_{i \in K} B_i y_i + \max_{\mathbf{x}_0} J_0^M(\mathbf{x}_0, \mathbf{N} \circ \mathbf{y}). \tag{19}
$$

Because initial allocation $\mathbf{x}_0$ is irreversible, it can only limit the firm's flexibility. It is easy to see from (18) that $J_0^M(\mathbf{x}_0, \mathbf{N})$ is decreasing in $\mathbf{x}_0$, hence minimizing $\mathbf{x}_0$ is optimal for Problem (19). This common industry practice, known as *gapping*, is practically implemented by overstating performance ratings (i.e. overselling ξ projections). Technically, this allows to set $\mathbf{x}_0 = \mathbf{0}$ wlog, and optimize Problem (18) for make-goods allocation only. In reality, however, excessive gapping can negatively affect client relationships on the long run.

5.1. Aggregate Model and Structural Properties

We first show that the multi-client dynamic model reduces to solving a much simpler model that aggregates all upfront clients into a single upfront mega-client, resulting in a dynamic program with a one-dimensional state variable. This allows us to easily characterize the structure of the optimal profit and make-goods allocation. Consider the following aggregate model for make-goods allocation:

$$
\begin{aligned}
J_t(N_t) = & \max_{x_0 \le x_t \le Q} && V_t(x_t, N_t) \\
\text{where} \quad & V_t(x_t, N_t) = && \pi(x_t) + \mathbb{E} J_{t+1}(N_t - \xi x_t) \\
\text{and} \quad & J_T(N_T) = && -bN_T^+.
\end{aligned}
\tag{20}
$$

Proposition 2 extends under the full make-goods allocation recourse as follows.

PROPOSITION 5. *The optimal solution to the client contracting Problem (19) with make-good recourse indicates that clients should be accepted in decreasing order of CPM. Moreover, the dynamic-make goods multi-client allocation Problem (18) is equivalent to the corresponding aggregate Problem (20) with target performance $N_t = \mathbf{1}'\mathbf{N}_t^+$, in that: (1) the value functions of the two problems are equal, $J_t^M(\mathbf{N}_t) = J_t(N_t)$, and (2) the total optimal make-goods allocation under model (18) equals the optimal make-goods allocation under the aggregate model (20), $\mathbf{1}'\mathbf{x}_t^* = x_t^*$. Furthermore, in each period, make-goods $x_{i,t}^*$ are optimally allocated to clients in proportion to their remaining performance targets $N_{i,t}^+$, i.e. by balancing GRP allocation.*

In conclusion, the following procedure for media planning is optimal based on model (19):

1. Serve clients in decreasing CPM order;

2. Allocate minimal initial inventory to clients upfront (gapping);

3. Obtain total make-goods allocation by solving Problem (20) with aggregate target performance $N_t = \mathbf{1}'\mathbf{N}_t^+$; its value function represents the expected year-end profit (net of client budgets).

4. Allocate make-goods $x^*_{i,t}$ to clients in proportion to their remaining performance targets $N^+_{i,t}$, i.e. by balancing GRP allocation.

In order to describe the multiple client solution, it is thus sufficient to characterize the optimal revenue and allocation corresponding to the aggregate model. Our next results rely on a standard comparative statics result, *Topkis' Lemma*, stating that if $g(x,y)$ has increasing (decreasing) differences then $x^*(y) = \max \operatorname{argmax}_{x\in X} g(x,y)$ is increasing (decreasing) in y (Topkis 1998). Recall that a bivariate function $g(x,y)$ has increasing (decreasing) differences in (x,y), if for all $x \le x'$, $g(x',y) - g(x,y)$ is increasing (decreasing) in y.[19]

The next result characterizes relevant structural properties of the BC's expected profit, formalizing the following statements: Expected profit increases with achieved performance, but its marginal value decreases. Moreover, the value of allocating an extra make-good decreases with achieved performance, and there is a diminishing marginal rate of substitution between current make-goods allocation and achieved performance $(N - N_t)$. Finally, the marginal value of airing an additional make-good (or lowering the performance target by one unit) is higher later in the horizon.

LEMMA 2. *The value function has the following properties: (a) J_t is decreasing and concave in N_t; (b) V_t is jointly concave and has increasing differences in (x_t, N_t); (c) $V_t(x,N) = V(x,N,t)$ has increasing differences in (x,t); (d) V_t has increasing differences in $(x,-N)$ and t.*

A consequence of Topkis' Lemma and Lemma 2(c), the next result shows that, at any point in time, the better the achieved performance, the lower the make-goods allocation, all else equal. To achieve the same given level of performance, more make-goods need to be aired later in the horizon. This is because there are less opportunities to make up for under-performance in the future.

PROPOSITION 6. *The optimal make-goods allocation $x^*_t(N)$ is increasing in the remaining performance target N at any time t, and decreasing in the remaining horizon $T - t$ for any $N \ge 0$, all else equal.*

[19] In two dimensions, increasing differences is equivalent to supermodularity. If X is a set of integers, monotone differences amounts to monotonicity of $g(x+1,y) - g(x,y)$.

In particular, there exist threshold levels $\bar{N}_t$ so that additional make-goods are aired at time t only if the remaining target exceeds the current threshold $N_t > \bar{N}_t$. These thresholds are increasing with achieved performance $N - N_t$, and over time, all else equal.

5.2. Extensions. Irreversible Allocation

This section describes extensions of the dynamic make-goods allocation model, illustrating the robustness and generality of the results presented in the previous section. We next present an alternative, irreversible allocation model, motivated by business practices in certain markets. Moreover, we note here that all the results in this and the previous subsection extend for Markovian audience processes. Furthermore, all comparative statics results extend under a general performance measure $\Psi(x, \xi)$ that is increasing and concave in x (accounting for repetition wearout).

Irreversible Allocation. Depending on the client's bargaining power, in certain markets make-goods allocation can also be irreversible (e.g. one additional P&G ad will be aired in *Friends* every week until the end of the season). Technically, if x_t is the total number of slots dedicated to a client at time t, irreversible allocation means that $x_{t+1} \geq x_t$, and effective available capacity at time $t+1$ is $Q - x_t$.[20] In practice, sellers do not necessarily resent this limited flexibility, as it often simplifies their decision making task and induces a more uniform allocation. This intuition is supported by our numerical results in Section 5.3.

The aggregate objective is $\max J_0(N)$, with the value function is given by the recursive Bellman equation:[21]

$$\begin{aligned} J_t(x_t, N_t) &= \max_{x_t \leq x_{t+1} \leq Q} \quad V_t(x_{t+1}, N_t) \\ \text{where} \quad V_t(x_{t+1}, N_t) &= \pi(x_{t+1}) + \mathbb{E} J_{t+1}(x_{t+1}, N_t - \xi x_{t+1}) \\ \text{and} \quad J_T(x_T, N_T) &= -b N_T^+. \end{aligned} \tag{21}$$

Given a pre-committed allocation x and a remaining target N, the optimal allocation policy is $x_t^* = x_t^*(x, N) = \arg\max_{x_t \in [x,Q]} V_{t-1}(x_{t+1}, N)$; the unconstrained optimum is denoted by $\bar{x}_t = \bar{x}_t(N) = \arg\max_y V_{t-1}(y, N)$.

Proposition 7. *All the results of Section 5.1 under reversible allocation extend to the irreversible allocation case. Moreover, the optimal irreversible allocation policy is given by* $x_t^*(x, N) = \max(x, \min(\bar{x}_t, Q))$, *which is also increasing in the previously committed allocation* x.

The result indicates that the BC should schedule no additional make goods unless the amount already committed x_t falls below a certain level, $\bar{x}_{t+1}$. Equivalently, the optimal policy is characterized by threshold levels $\bar{N}_{t+1} = N_t - \bar{x}_{t+1}\mu$, so that additional make-goods are aired only if the

[20] A conceptually related stream of literature in manufacturing considers firms' capacity expansion strategies as irreversible investments (see e.g. Oksendal 2000).

[21] Again, gapping considerations suggest that initial upfront allocation should be minimized, so w.l.o.g. we set $x_0 = 0$.

remaining target exceeds the current threshold. These thresholds are increasing with past allocation x, achieved performance $N - N_t$, and over time, all else equal. Section 5.3 (Figure 1) illustrates these results numerically, contrasting them with the reversible allocation case.

Such structural properties, albeit intuitive, are typically difficult to preserve in a dynamic and stochastic setting. In particular, the monotonicity of the optimal allocation over time is a subtle result in the irreversible case, complicated by the state dependence of the action set $\mathcal{S}_x = [x, Q]$.[22]

5.3. Heuristics and Numerical Results

This section provides heuristics for dynamic make-goods allocation, and analyses their performance compared to the optimal policy. Numerical results are illustrated for an isoelastic scatter market demand function $d(p) = ap^{-\eta}$, with elasticity $\eta > 1$. This leads to a profit model $\pi(x) = ap^{1-\eta} = p_0(Q - x)^{1-1/\eta}$, where $p_0 = a^{1/\eta}$ corresponds to the profit from one unit inventory available to the scatter market. We set $p_0 = 5$ (e.g. in tens of thousand of dollars), and $\eta = 1.5$. Audience is modeled as truncated normal distribution with $\mu = 4$ and $\sigma = 2$, measured in millions of eyeballs (this would correspond to a Binomial model with market size $M = 8$ million and the viewing probability $p = .5$). Target performance N is also measured in millions (generally in the order of a few hundreds). Total ad space capacity Q is set to 30, corresponding to the number of 30 second spots in a 15 minute break. We report results for penalty levels b of 10 and 70, corresponding approximately to 70% and 90% service levels (see Proposition 8(a) of Section 6.1 and Figure 3). Penalties b have the same order of magnitude as p_0, so for an ad price of $p_0 = \$50,000$, the penalties considered are $b = \$100$, respectively \$700 per thousand eyeballs. As expected, this value is much higher than scatter market CPM, which in our case is at most $p_0/\mu = \$12.5$. Our extensive numerical studies, for a wide range of meaningful numerical values, indicate that the results presented here are representative and robust.

Figure 1 supports our theoretical findings by illustrating the behavior of the optimal allocation policy and expected profits with respect to time and remaining target, for both the reversible and irreversible models. As predicted, expected profit is decreasing and concave in the remaining target N, and decreasing with respect to remaining horizon length $T - t$. Capacity allocation x is higher, the higher the remaining target N, and the shorter the remaining horizon. Figure 1 also illustrates the suboptimality of irreversible allocation, which postpones allocation to later periods in order to avoid being "stuck" early with too high commitments to the upfront.

[22] For instance, this excludes the use of Smith and McCardle's (2002) meta-theorems for structural properties of dynamic programs with action independent sets (e.g. their Proposition 5, p.806).

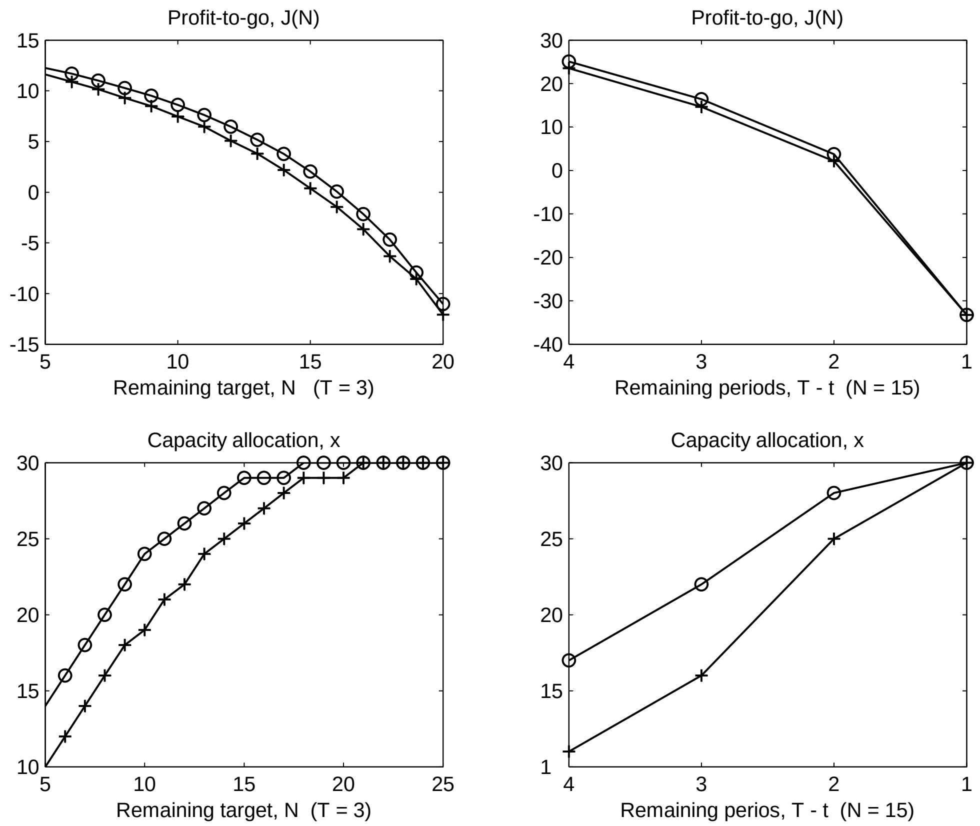

Figure 1 **Optimal profit-to-go and capacity allocation for reversible (—○—○) and irreversible (—+—) models as a function of remaining target** N **and remaining time** $T-t$**;** $b=10$**.**

The practical inconvenience of solving the dynamic program to optimality (due to the curse of dimensionality) motivates us to study several intuitive and efficient heuristics for make-goods allocation, described next

The *Certainty Equivalent Control* (CEC) heuristic (see e.g. Bertsekas 2000) is a typical approach in approximate dynamic programming that obtains a deterministic approximation of the profit-to-go by replacing the underlying randomness by its expectation. In our setting, this corresponds to approximating the audience in all remaining periods by its expected value $\mathbb{E}\xi = \mu$. Effectively, in each period the following deterministic concave problem is solved:

$$\begin{aligned} \max \quad & \sum_{i=t}^{T} \pi(x_i) - bZ \\ \text{s.t.} \quad & Z \geq N_t - (\sum_{i=t}^{T} x_i)\mu \end{aligned}$$

$$Z \geq 0 \text{ and } 0 \leq x_i \leq Q, \quad t \leq i \leq T$$

Concavity of π implies that, at an optimal solution, the allocation is uniform and equal to $x_t^* = \max\{\arg\max(\pi(x) + xb\mu), \frac{N_t}{(T-t+1)\mu}\}$. This is the amount allocated by the CEC policy in period t, calculated *as if* the firm would set a uniform allocation across remaining periods. In each period, however, the policy re-solves the corresponding deterministic approximation, after updating the remaining target based on realized audience. So effectively allocation is not uniform over time.

We also propose a *minimal postponement* policy, which prioritizes allocation to the upfront market, while approximating future audience by its mean. Only once the client target is met, the policy starts selling to the scatter market. Specifically, it sets $x_t = \min(Q, N_t/\mu)$, as long as $N_t > 0$. One can define an analogous *maximal postponement* policy, but this turns out to perform relatively poorly, hence is not reported.

The *myopic* policy reduces the visibility of the decision maker to a shorter horizon, by introducing *sub*-targets for corresponding *sub*-horizons. For clarity, we consider the extreme myopic situation where each period is a sub-horizon and the target N is initially divided uniformly across periods. The myopic policy allocates a constant amount of inventory each period, corresponding to a per period target N/T, as long as this does not exceed the actual remaining target. Concretely, in each period t, the myopic allocation x_{myopic} solves the static model (A) with target $\tilde{N}_t = \min(N/T, N_t)$, where N_t is the total actual remaining target.

Finally, the *static* policy provides the corresponding optimal open-loop policy, deciding allocations for each period before audience realizations are revealed. This model is analogous to the multi-stage model (CT) of Section 3.3 with non-linear scatter market profit. Proposition 3 (b) insures that the static policy allocates the same amount of make-goods x_{static} in each period. Specifically, the static solution solves model (A) with target N and cumulative audience distribution equal to the convolution $\xi^{(T)} = \xi_1 + \cdots + \xi_T$, where $\xi_i, \ i = 1, \ldots, T$ are i.i.d. with the same distribution as ξ, i.e. x_{static} solves:

$$\max_{x \in [0,Q]} T\pi(x) - b\mathbb{E}\left[N - x\xi^{(T)}\right]^+. \tag{22}$$

By design, the static policy is implicitly also a valid heuristic for the irreversible model.

Figure 2 compares numerically the performance of these policies in terms of total expected profits and allocation (calculated by Monte Carlo simulation) at the beginning of a four quarter horizon $T = 4$. The CEC policy is omitted from this figure because it is systematically outperformed by the others (essentially because its allocation to the upfront is too low). Myopic policies achieved

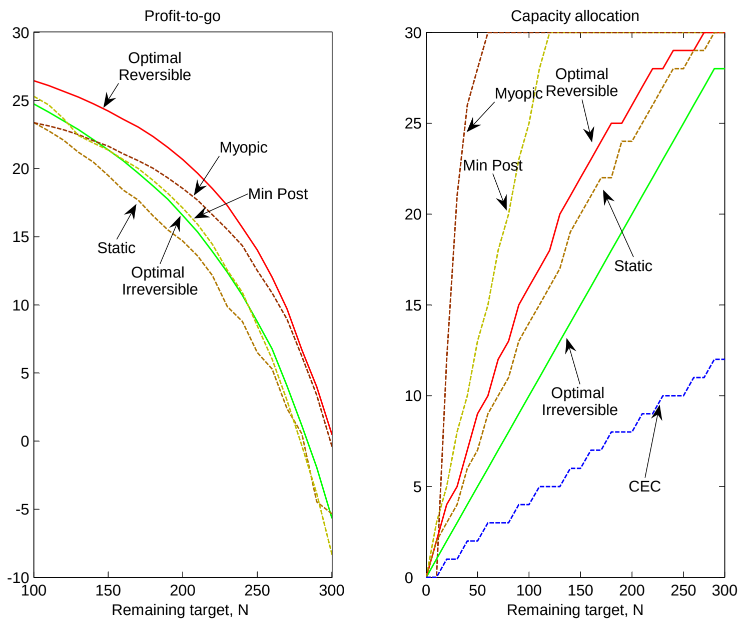

Figure 2 **The value function and allocation policy at time $t = 1$ as functions of the target N, for $T = 4$ and $b = 70$.**

the best value function approximation, with a relative error with respect to the optimal reversible policy in the order of 0.06 (relative error is defined as the percentage optimality gap, here $(J_{rev} - J_{Myopic})/J_{rev}$).[23] Moreover, the myopic policy improves significantly as the performance target N increases. For low performance targets, the myopic policy is outperformed by the minimum postponement. Because both policies favor the upfront market, their performance improves as the penalty value increases.[24] The static policy is the only heuristic among the four considered here that is also feasible for the irreversible allocation settings, achieving an average relative error in the order of .09.[22]

The second graph in Figure 2 shows the corresponding optimal allocation of each heuristic at the beginning of a four period horizon. Both myopic and minimum postponement policies neglect the opportunity of correcting performance in future allocations, hence predict higher upfront allocations. The static allocation policy comes closest to the optimal policy, even though its corresponding revenue is outperformed by both minimal postponement and myopic policies. This is due to the

[23] Relative *profit* performance is actually much better, because it should also consider the upfront budget B, omitted by our value function calculations. To give an idea of the magnitude of B, in 2005, the six biggest networks in the US alone collected over \$9 billion in upfront revenues, that is $1/7$ of the \$63 billion TV advertising market, according to TNS. So, accounting for B, relative profit performances would be at most $6/7 = 86\%$ of reported values.

[24] Indeed, their performance is worse than the static policy only for extremely low penalties $b \leq 2$, corresponding to impractical service levels in the order of 30%.

asymmetry of the value function. Our numerical results suggest a systematic ordering of these policies, as follows $x_{irrev} \le x_{static} \le x_{rev} \le x_{MinPost}$. We provide an intuitive argument for this. Because the irreversible policy can (only) increase initial allocation in the future, it is expected to set a lower initial allocation than the static one. Relative to the reversible regime, the static policy does not have the opportunity to decrease make-goods allocation over time, so it sets a lower first stage allocation (to avoid unrewarded over-performance). By design, the minimum postponement policy starts by allocating maximum capacity to the upfront (based on an average audience) $x_{MinPost} = \min(Q, N/\mu)$, which is expected to exceed the optimal reversible allocation. We also often observe numerically that $x_{MinPost} \le x_{myopic}$, which is obvious for reasonably large targets that exceed average one-period ratings capacity, $N > Q\mu$.

We conclude that myopic policies based on the static upfront allocation model (A) presented in Section 3 provide a surprisingly good approximation for the complex dynamic make-goods optimization problem. In addition, intuitive minimal postponement and static policies also provide good approximations for reversible, respectively irreversible regimes.

6. Extensions. Service Constrained Models and Long Term Strategies

Our models assumed a linear penalty per unit of unmet performance. In practice, under-performance penalties are often implicit, as they must indirectly account for retention factors such as loss of goodwill, potential loss of client etc. This section shows that the key results obtained under the penalty framework are preserved under alternative service constrained models. We also discuss service differentiation and long term strategies.

6.1. Service Models

Suppose that the BC wants to insure that the probability of meeting the contracted target performance exceeds a preset type 1 service level $1-\epsilon$. The following aggregate service model obtains:

$$(S) \qquad \begin{aligned} &\min_{0\le x\le Q} \; px \\ &\text{s.t. } \mathbb{P}(x\xi \le N) \le \epsilon. \end{aligned} \tag{23}$$

This problem is feasible provided that $N/F^{-1}(\epsilon) \le Q$. The resulting allocation is $x^* = N/F^{-1}(\epsilon)$, i.e. the optimal GRP allocation satisfies the critical fractile condition $w^* = F^{-1}(\epsilon)$. Note, this is actually independent of the scatter market profit model.

Imposing a uniform strategic service level $1-\epsilon$ across clients leads to the following service-counterpart of the multi-client contracting and inventory planning penalty model (7):

$$\begin{aligned} \max\; & p(Q - \textstyle\sum_i x_i) + \sum_i B_i y_i \\ & y_i(\mathbb{P}(x_i\xi \le N_i y_i) - \epsilon) \le 0, \quad i = 1,\dots,K \\ & \textstyle\sum_i x_i \le Q, x_i \ge 0 \\ & y_i \in Y = [0,1]. \end{aligned} \tag{24}$$

The following result shows that the penalty and service models lead to consistent predictions:

PROPOSITION 8. *(a) For the same contracted target N, the optimal solution of the penalty model (A) exceeds that of the service model (S) if and only if $G^{-1}(p/b) \le F^{-1}(\epsilon)$. In particular, the aggregate service and penalty models are consistent, i.e. result in the same optimal allocation x^*, if and only if $G^{-1}(p/b) = F^{-1}(\epsilon)$.*[25]

(b) The optimal solution of the multi-client service model (24) is to serve clients in decreasing order of CPM C_i, as long as this exceeds $p/F^{-1}(\epsilon)$ (i.e. clients are profitable), and capacity is still available. Furthermore, it is optimal to equate GRP allocation across clients to the optimal aggregate upfront GRP allocation $w^* = F^{-1}(\epsilon)$.

Figure 3 illustrates the correspondence between service and penalty levels identified in part (a).

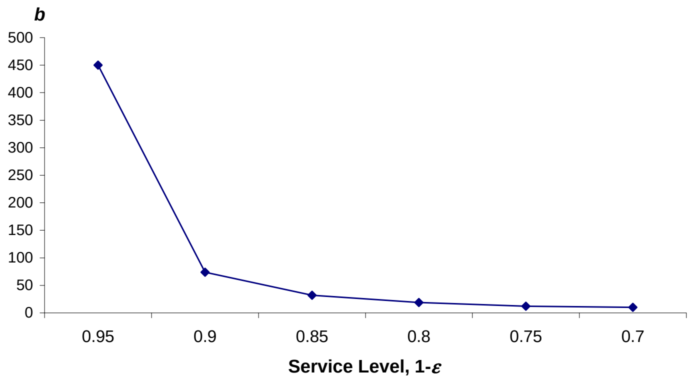

Figure 3 **Penalty b vs. service level $1-\epsilon$; ξ is truncated $N(4,2)$ and $p=5$.**

Under the condition of Proposition 8(a), the service and penalty model prescriptions differ only in the threshold for client acceptability, $p/F^{-1}(\epsilon) \neq bF(G^{-1}(p/b)) = b\epsilon$; the discrepancy is due to different (marginal) profits in the two models. The change of measure transformation introduced in Section 4.1 sheds light on this result, revealing a one to one relation between penalty and service constrained models for media planning. To see this, recall the inventory reformulation (13) of model (A), with underage cost p and overage cost $b\mu - p$. The performance of this model (under the measure ζ) is $\epsilon = p/(p+\mu b - p) = p/(\mu b)$, i.e. the ratio of the average opportunity cost of reaching one eyeball (scatter market CPM) p/μ and the penalty cost b of missing it. Because ζ dominates

[25] In this case, feasibility of the service model implies that the unconstrained solution is optimal for the penalty model.

ξ, the service constrained model (S) allocates less than the penalty model (A) with $b = p/(\epsilon\mu)$, as confirmed by Proposition 8 (a). Model (S) corresponds to an inventory model with inventory-unit demand N/ξ, and conversely, model (A) corresponds to a service model (S) with ξ replaced by ζ.

The one-stage service constrained model (S) can also be extended to handle dynamic make-goods allocation (the Bellman recursions only differ in the terminal value function). This yields the same structural results as those obtained for penalty models in Section 5.

All our insights remain valid under type 2 service models, which impose a rigid (strategic) bound $\delta > 0$ on the expected fraction of unmet performance: $\mathbb{E}[N_i - x_i\xi]^+ \leq \delta N_i$. The results of penalty and type 1 service models are reproduced with optimal GRP allocation $w^* = L^{-1}(\delta)$, where $L(w) = \mathbb{E}[1 - \xi/w]^+ = F(w) - G(w)/w$. Such models provide a realistic formalization of the approach adopted in practice, where client constraints are satisfied within a range controlled by the sales manager (see Bollapragada et al. 2002, p.54).

6.2. Service Differentiation and Long Term Profitability

In practice, the BC typically prioritizes upfront market requests in order of their profitability and importance to the company:"Clients that usually pay high premiums and spend a lot of money with NBC are given high priority" (Bollapragada et al. 2002, p. 49). This is consistent with our static results, for a *long-term* profit optimization model that optimizes *expected future revenues* from accepted clients, instead of present ones. Expected future revenue from customer i can be estimated by weighing budget values B_i by expected client lifetime L_i. Our model in this case suggests prioritizing clients in order of lifetime adjusted CPM, C_iL_i, a proxy for client's value, or importance to the company.

Our models assumed a strategic service level $1-\epsilon$ for all clients (in the penalty case, this is captured by the unique penalty b). Differentiating service levels $1-\epsilon_i$ across clients at the strategic contracting phase (24) results in the recommendation to prioritize clients according to service-adjusted CPM $C_iF^{-1}(\epsilon_i)$. In particular, among clients of similar immediate profitability C_i, the BC will favor easier to fulfil contracts (hence potentially less important clients). This motivates the need for more complex models that (optimally) set different service levels $1-\epsilon_i$ for each client, while capturing the endogenous impact of service level on customer lifetime value $L_i(\epsilon_i)$. Such recourse models are relevant, but beyond the scope of the current paper.

7. Conclusions

This paper provides stylized models for media revenue management in the presence of audience uncertainty. Our results provide several levels of decision support for broadcasting companies to optimize revenues from advertising space. We devise a simple procedure for accepting upfront client contracts and estimating their overall inventory requirements. The impact of audience uncertainty on these decisions is also assessed. Operationally, we indicate how much inventory should be initially committed to upfront clients before the start of the season, with the provision that additional make-goods can be aired subsequently. Finally, we make recommendations for dynamic make-goods allocation during the scatter market. We provide structural results and efficient heuristics to support high-level managers and account executives in this complex decision process. Our results are robust, in that they hold under general models of audience uncertainty and scatter market profit, as well as under alternative service constrained models.

Our models can also be useful in other settings where the value of supply is uncertain, and/or the firm serves dual markets. For instance, in manufacturing with random yield, a firm with limited resources and uncertain production rate (e.g. machine or workforce reliability) honors both key accounts with long term contracts and small clients with opportunistic contracts (alternatively, our scatter market opportunity cost corresponds to direct production cost). Similarly, non-profit organizations (NPO) afford to sustain pro bono service to a mission market by often offering similar paid services to distinct markets, that share the same limited facilities (e.g. hospitals) powered by uncertain resources (e.g. volunteers). In this case, the mission market corresponds to our upfront market and the paying market to our scatter market. de Vericourt and Lobo (2006) investigate revenue management for NPOs.

There are multiple facets of the media revenue management problem that this paper leaves to be explored, including pricing, scheduling and contract design. A static deterministic integer programming model for spot scheduling is proposed by Bollapragada and Garbiras (2004). Contract design is an interesting area, where the industry is quite sophisticated, managing a variety of flexible products, such as option-cutbacks (revised planning due to client's budget cuts) or callable products (lower rates with the option of recalling the slot for a higher paying customer). The latter, and their value for standard B2C revenue management are investigated in Gallego (2004).

Finally, it is an irreversible fact that audience is moving online, where inventory is more complex, dynamic and customizable, leading to different contractual terms (e.g. pay per click vs CPM – pay per view). Araman and Fridgeirsdottir (2006) investigate a queuing model for online advertising revenue management.

Acknowledgments

The authors would like to thank Michael Chigirinskiy of Invision Inc. and Monique Garnier and Antoine Taconnet of former Peaktime (now eTelmar) for insightful discussions about practical aspects of media planning and revenue management.

References

Araman, V.F., K. Fridgeirsdottir 2006. Online advertisement: A revenue management approach. *Working paper.*

Araman, V.F., O. Ozer 2005. Capacity and inventory management in the presence of a long-term channel and a spot market, *Working paper, New York University.*

Bassok, Y., W. J. Hopp, M. Rohatgi. 2002. A simple heuristic for the service constrained random yield problem. *IIE Transactions* **34** 479-487.

Bertsekas, D. 1995. *Nonlinear Programming.* Athena Scientific.

Bertsekas, D. 2000. *Dynamic Programming and Optimal Control.* Athena Scientific.

Bollapragada, S., H. Cheng, M. Phillips, M. Garbiras. 2002. NBC's optimization systems increase revenues and productivity. *Interfaces* **32** (1) 47-60.

Bollapragada, S., H. Mallik, M. 2006. Managing the inventory of commercial airtimes in broadcast television. *Working paper.*

Bollapragada, S., M. Garbiras. 2004. Scheduling commercials on broadcast television. *Operations Research* **53** (3) 337-345.

Bollapragada, S., T.E. Morton. 1999. Myopic heuristics for the random yield problem. *Operations Research* **47** (5) 713-722.

de Vericourt, F., M. S. Lobo. 2006. Resource and Revenue Management in Nonprofit Operations. *Fuqua Business School working paper.*

Gallego, G., S.G. Kou, R. Phillips. 2004. Revenue Management of Callable Products. *working paper*

Grosfeld-Nir, A., Y. Gerchak. 2004. Multiple lotsizing in production to order with random yields: Review of recent advances. *Annals of Operations Research* **126** 43-69.

Gupta, D., W. L. Cooper. 2005. Stochastic comparisons in production yield management. *Operations Research* **53** (2) 377-384.

Kanetkar, V., C.B. Weinberg, D.L. Weiss. 1992. Price Sensitivity and Television Advertising Exposures: Some Empirical Findings. *Marketing Science* **11** (4) 359-371.

Kleindorfer, P and D.J. Wu. 2003. Integrating long-term and short-term contracting via business-to-business exchanges for capital-intensive industries. *Management Science* **49** (11) 1597-1615.

Lodish, L.M., M. Abraham, S. Kalmenson, J. Livelsberger, B. Lubetkin, B. Richardson, M.E. Stevens. 1995. How TV Advertising Works: A Meta-Analysis of 389 Real World Cable TV Advertising Experiments. *Journal of Marketing Research* **32** 125-139.

Müller, A., D. Stoyan. 2002. *Comparison Methods for Stochastic Models and Risks.* Wiley Publishers.

Porteus, E. L. 2000. *Foundations of Stochastic Inventory Theory.* Stanford Business Books.

Oksendal, A. 2000. Irreversible investment problems. *Finance Stochastics* **4** 223-250.

Savin,S., N. Gans, M. Cohen, and Z. Katalan. 2005. Capacity Management in Rental Businesses with Two Customer Bases, *Operations Research* **53** 617-631

Smith, J. E., K.E. McCardle 2002. Structural Properties of Stochastic Dynamic Programs, *Operations Research* **50** 796-809.

Sepehri, M., E. A. Silver, C. New. 1986. A Heuristic for multiple lotsizing for an order under variable yield. *IIE Transactions* **18** (1) 63-69.

Talluri, T. T., G. J. Van Ryzin. 2004. *The Theory and Practice of Revenue Management.* Kluwers Academic Publishers.

Topkis, D. M. 1998. *Supermodularity and Complementarity.* Princeton University Press.

Yano, C., H. L. Lee. 1995. Lot sizing and random yields: A review. *Oper. Res.* **43**(2) 311-334.

Wu, D. J., P. Kleindorfer. 2003. Competitive options, supply contracts and electronic markets. *Management Science* **51** (3) 452-466.

Online Technical Appendix to "Stochastic Revenue Management Models for Media Broadcasting"

Appendix A: Proof of Proposition 3 (Multi-stage planning)

Part (a)related to model (MT) follows directly from Proposition 1. Our proof of part (b) relies on the following lemma (see Theorem 8.2.3 in Müller and Stoyan 2002), concerning the problem:

$$(U) \quad \min_{\mathbf{1}'\mathbf{y}=q,\, \mathbf{y}\geq 0} \mathbb{E}u(\mathbf{y}'\omega). \tag{25}$$

LEMMA 3. *Suppose that u is a decreasing convex function and ω_i are exchangeable. The optimal solution $\mathbf{y}^* \in \mathbb{R}^n$ to Problem (U) above satisfies $y_i^* = q/n$ for all i.*

Problem (CT) can be written as

$$(P_Q) \quad \min_{\mathbf{x}\in[\mathbf{0},\mathbf{Q}]} p\mathbf{1}'\mathbf{x} + b\mathbb{E}[N - \mathbf{x}'\boldsymbol{\xi}] = \min_{q\geq 0} pq + bZ(q), \tag{26}$$

$$\text{where} \quad Z(q) = \min_{\substack{\mathbf{x}\in[\mathbf{0},\mathbf{Q}] \\ \mathbf{1}'\mathbf{x}=q.}} \mathbb{E}[N - \mathbf{x}'\boldsymbol{\xi}]^+ \tag{27}$$

We prove that for any q, the solution $\mathbf{x}^*(q)$ of Problem (27) has the desired structure, which implies the result for the solution of the original problem $\mathbf{x}^*(q^*) = \mathbf{x}^*$.

Problem (27) without capacity constraints has the structure required by Lemma 3. So, in particular, if the optimal solution $\bar{\mathbf{x}}$ of this problem satisfies $\bar{\mathbf{x}} \leq \mathbf{Q}$, then it is optimal, so the result of Lemma 3 holds. Otherwise, consider an optimal solution $\mathbf{x}^* = \mathbf{x}^*(q)$ to Problem (CT). Consider the *reduced problem* obtained by fixing the variables that are equal to their upper boundary ($\mathbf{x}_I = \mathbf{x}_I^* = \mathbf{Q}_I$) in the given solution, and optimizing the objective over the remaining subset of variables $\mathbf{x}_J$, in absence of capacity constraints $\mathbf{Q}_J$. For a vector $\mathbf{y} \in \mathbb{R}^k$, we denoted its projection on the $L \subseteq \{1,\ldots,K\}$ coordinate set by $\mathbf{y}_L = (y_l)_{l\in L}$. Letting $\eta_J = N - \mathbf{x}_I^*\boldsymbol{\xi}_I$ (random) and $q_J = q - \mathbf{1}'\mathbf{Q}_I$, the reduced problem is:

$$\min_{\substack{\mathbf{v}_J\geq \mathbf{0} \\ \mathbf{1}'\mathbf{x}_J = q_J.}} \mathbb{E}[\eta_J - \mathbf{x}_J'\boldsymbol{\xi}_J]^+ \tag{28}$$

This reduced problem has the structure required by Lemma 3, hence its solution $\bar{\mathbf{x}}_J$ satisfies the corresponding properties. Moreover $\mathbf{x}^*$ given by $x_J^* = \bar{\mathbf{x}}_J, \mathbf{x}_I^* = \mathbf{Q}_I$ is optimal to Problem (CT) (see e.g. Proposition 3.4.2 in Bertsekas, 1995). This proves the desired results. □

Appendix B: Proof of Proposition 4 (Audience sensitivity)

(a) The first order condition can be restated as $H_{\zeta(z)}(N/x) - bC(z) = 0$. The assumptions imply that the left hand side is decreasing in z, hence so is $x^*(z)$.

(b) Lemma 1 suggests that the optimal (unconstrained) allocation solves:

$$G(N/x,z) = p(z)/b, \text{ where } G(N/x,z) = \mathbb{E}[\xi(z);\xi(z)\leq N/x] = \int_0^{N/x} yf(y,z)dy. \tag{29}$$

Here $f(y,z) = f_0(y - \mu(z))$ denotes the density function of $\xi(z)$. This is decreasing in z for $y \leq \mu(z)$, because f_0 is unimodal with non-negative mode. Thus, $G(N/x,z)$ is decreasing in z for $x \geq N/\mu(z)$.

The condition on b implies $x^*(z) > N/\mu(z)$. Monotonicity of $x^*(z)$ follows because $p(z)$ is increasing in z and $G(N/x,z)$ is decreasing x, and decreasing in z for $x \geq N/\mu(z)$. □

Appendix C: Proofs for Section 5.1 (Reversible make-goods allocation)

C.1. Proof of Proposition 5:

The proof is by induction. The base case follows the same lines as Proposition 1, but for general $\pi(x)$. We provide it here for completeness. Let $N = \mathbf{1}'\mathbf{N}^+$, and observe that

$$J^M_{T-1}(\mathbf{N}) = \max_{\mathbf{0}\le\mathbf{x};\mathbf{1}'\mathbf{x}\le Q} \pi(\mathbf{1}'\mathbf{x}) - b\mathbb{E}[\mathbf{1}'(\mathbf{N}-\xi\mathbf{x})^+] \le \max_{0\le X\le Q} \pi(X) - b\mathbb{E}[(N-\xi X)^+] = J_{T-1}(N). \tag{30}$$

The inequality holds by taking $X = \mathbf{1}'\mathbf{x}$ in the right hand side, and using the fact that $\sum a_i^+ \ge (\sum a_i)^+$.

Consider the optimal allocation X^* that optimizes the right hand side above, and let $\mathbf{x}$ so that $x_i/N_i = X^*/N$, for all i with $N_i > 0$. This satisfies $\sum x_i = X^*$, and $\mathbf{1}'(\mathbf{N}-a\mathbf{x})^+ = (N-aX)^+$, for any $a \ge 0$, so $\mathbf{x}$ achieves (30) with equality. This shows that $\mathbf{x}$ defined by $x_i/N_i = X^*/N$ is optimal to the multi-client model, and $J^M_{T-1}(\mathbf{N}) = J_{T-1}(N)$.

Now, assuming the result is true for $t+1$, we show that it also holds for t. Assume wlog that $\mathbf{N} \ge 0$; otherwise replace $\mathbf{N}$ by its positive part, because $J^M_t(\mathbf{N}) = J^M_t(\mathbf{N}^+)$. For $X = \mathbf{1}'\mathbf{x}$ we have:

$$\begin{aligned} V^M_t(\mathbf{x},\mathbf{N}) &= \pi(\mathbf{1}'\mathbf{x}) + \mathbb{E}J^M_{t+1}((\mathbf{N}-\xi\mathbf{x})^+) \\ &= \pi(X) + \mathbb{E}J_{t+1}(\mathbf{1}'(\mathbf{N}-\xi\mathbf{x})^+) \quad \text{(induction step)} \\ &\le \pi(X) + \mathbb{E}J_{t+1}((N-\xi X)^+). \quad \text{(monotonicity of } J_t\text{)} \end{aligned} \tag{31}$$

Therefore,

$$\begin{aligned} J^M_t(\mathbf{N}) &= \max_{\mathbf{0}\le\mathbf{x};\mathbf{1}'\mathbf{x}\le Q} V^M_t(\mathbf{x},\mathbf{N}) \\ &= \max_{0\le X\le Q} \pi(X) + \mathbb{E}J_{t+1}(\mathbf{1}'(\mathbf{N}-\xi\mathbf{x})^+) \\ &\le \max_{0\le X\le Q} \pi(X) + \mathbb{E}J_{t+1}((N-\xi X)^+) \\ &= J_t(N). \end{aligned} \tag{32}$$

Consider the optimal allocation X^* that optimizes $J_t(N)$ above, and let $\mathbf{x}$ so that $x_i/N_i = X^*/N$, for all i with $N_i > 0$. This is feasible and achieves (32) with equality, hence it is an optimal solution to the multi-client model, and the two value functions are equal.

Finally, the first statement of the proposition follows from the aggregation result, along the lines of proof of Proposition 2(a). □

C.2. Proof of Lemma 2:

The terminal value function J_T is decreasing and concave in N_T. Together with concavity of π_t, this implies that V_{T-1} is jointly concave and has increasing differences. By induction we obtain that V_t is jointly concave and has increasing differences in (x,N) and J_t is decreasing concave in N_t implying (a) and (b).

Parts (c) and (d) follow by writing $V_{t-1}(x,N) - V_t(x,N) = \mathbb{E}[J_t(N-x\xi) - J_{t+1}(N-x\xi)]$. Concavity (hence decreasing differences) of J_{t+1} implies that, for any realization z of ξ, the function inside this expectation:

$$\max_{x_0\le y\le Q} \pi(y) + \mathbb{E}[J_{t+1}(N-xz-y\xi) - J_{t+1}(N-xz)] \tag{33}$$

is increasing in $(N,-x)$. □

Appendix D: Proofs for Section 5.2 (Irreversible Allocation and Extensions)

D.1. Preliminary Technical Lemmas

We present a set of general technical lemmas that are used to prove the results in this section. The generality of the presentation allows to easily extend the proofs for general performance metrics $\Psi(x,\xi)$, validating the claims at the beginning of Section 5.2.[26]

LEMMA 4.

(a) f concave implies $g(x,y)=f(x+y)$ is concave and has decreasing differences.

(b) f concave and h increasing implies $g(x,y)=f(h(x)+y)$ has decreasing differences.

(c) f increasing concave and h concave implies $g(x,y)=f(h(x)+y)$ joint concave.

(d) If $f(x,y)$ has decreasing differences in (x,y) and is concave in y and $h(x)$ is increasing then $g(x,y)=f(x,h(x)+y)$ has decreasing differences in (x,y).

Analogous conditions for increasing differences obtain from the fact that $g(x,y)$ has increasing differences if and only if $g(-x,y)$ has decreasing differences if and only if $g(x,-y)$ has decreasing differences.

Proof: (a) Follows from the definitions.

(b) For $x\le x'$ and $y\le y'$, concavity of f implies:

$$\begin{aligned} g(x',y')-g(x',y) &= f(y'+h(x'))-f(y+h(x')) \\ &\le f(y'+h(x))-f(y+h(x)) = g(x,y')-g(x,y). \end{aligned}$$

(c) Denote $x_\alpha=\alpha x+(1-\alpha)x', y_\alpha=\alpha y+(1-\alpha)y'$. We have

$$g(x_\alpha,y_\alpha)=f(h(x_\alpha)+y_\alpha)\ge f(\alpha(h(x)+y)+(1-\alpha)(h(x')+y'))\ge \alpha f(h(x)+y)+(1-\alpha)f(h(x')+y'),$$

where the first inequality is by concavity of h and monotonicity of f, and the second by concavity of f.

(d) For $x\le x'$ and $y\le y'$ we have:

$$\begin{aligned} g(x',y')-g(x',y) &= f(x',y'+h(x'))-f(x',y+h(x')) \\ &\le f(x',y'+h(x))-f(x',y+h(x)) \\ &\le f(x,y'+h(x))-f(x,y+h(x)) \\ &= g(x,y')-g(x,y), \end{aligned}$$

where the first inequality follows by concavity (hence decreasing differences) of f in the second argument, and the second part by decreasing differences of f in (x,y).

□

The next lemma requires the following definition:

DEFINITION 1. The set function $x\mapsto\mathcal{S}_x$, or in short $\mathcal{S}_x$, is said to be:

(a) decreasing if $\mathcal{S}_{x'}\subseteq\mathcal{S}_x$ for $x\le x'$; increasing if $\mathcal{S}_{x'}\subseteq\mathcal{S}_x$ for $x\ge x'$.

(b) convex if $y_i\in\mathcal{S}_{x_i},\ i=1,2$ implies $\alpha y_1+(1-\alpha)y_2\in\mathcal{S}_{\alpha x_1+(1-\alpha)x_2}$ for all $\alpha\in[0,1]$.

[26] The proof for $\Psi(x,\xi)$ uses Lemma 4 b) and c). The induction in Proposition 7 should be conducted simultaneously for *all* increasing concave functions $\Psi(x,\xi)$, in order to achieve the last step of the proof.

LEMMA 5. *(a) The function* $g(x) = \max_{y \in \mathcal{S}_x} f(y)$ *is decreasing in* x *if* $\mathcal{S}_x$ *is decreasing in* x.

(b) The function $g(x) = \max_{\mathbf{y} \in \mathcal{S}_x} f(x,y)$ *is concave in* x *if* $f(x,y)$ *is jointly concave in* (x,y) *and* $\mathcal{S}_x$ *is convex in* x.

(c) The function $g(x,y) = \max_{v \in [x,Q]} f(v,y)$ *has decreasing (increasing) differences in* (x,y) *if* $f(v,y)$ *has decreasing (increasing) differences in* (v,y) *and is concave in* v.

Proof: (a) Trivial

(b) Given x_1, x_2 and $\alpha \in (0,1)$, denote $x_\alpha = \alpha x_1 + (1-\alpha)x_2$. We have

$$\alpha g(x_1) + (1-\alpha) g(x_2) = \alpha f(x_1, y_1^*) + (1-\alpha) f(x_2, y_2^*) \tag{34}$$

$$\leq f(x_\alpha, \alpha y_1^* + (1-\alpha) y_2^*) \tag{35}$$

$$\leq \max_{y \in \mathcal{S}_{x_\alpha}} f(x_\alpha, y) = g(x_\alpha) \tag{36}$$

where the first inequality holds by joint concavity of f, and the second uses the fact that $\alpha y_1^* + (1-\alpha) y_2^* \in \mathcal{S}_{x_\alpha}$ for $y_1^* \in \mathcal{S}_{x_1}$ and $y_2^* \in \mathcal{S}_{x_2}$ by convexity of the correspondence $\mathcal{S}_x$.

(c) We only prove the decreasing differences result here; the other is analogous. By concavity of f in v, we can write $\operatorname{argmax}_{x \leq v \leq Q} f(v,y) = \max(x, \bar{x}_y)$, where $\bar{x}_y = \operatorname{argmax}_{0 \leq v \leq Q} f(v,y)$. By Topkis' Lemma, $\bar{x}_y$ is decreasing in y.

Decreasing differences of g amounts to proving for any $x \leq x'$ and $y \leq y'$:

$$\begin{aligned} g(x',y') + g(x,y) &= \max_{x' \leq v' \leq Q} f(v',y') + \max_{x \leq v \leq Q} f(v,y) \\ &= f(\max(x', \bar{x}_{y'}), y') + f(\max(x, \bar{x}_y), y) \\ &\leq f(\max(x, \bar{x}_{y'}), y') + f(\max(x', \bar{x}_y), y) \\ &= \max_{x \leq v \leq Q} f(v,y') + \max_{x' \leq v' \leq Q} f(v',y) = g(x,y') + g(x',y). \end{aligned} \tag{37}$$

If $X = \max(x, \bar{x}_y) \leq \max(x', \bar{x}_{y'}) = X'$, then (37) follows by decreasing differences of f:

$$f(X',y') + f(X,y) \leq f(X,y') + f(X',y) \leq \max_{x \leq v \leq Q} f(v,y') + \max_{x' \leq v' \leq Q} f(v',y).$$

It remains to show it for $X' \leq X$. Because $\bar{x}_{y'} \leq \bar{x}_y$ (from $y' \geq y$) and $x' \geq x$, we obtain $x \leq x' \leq \bar{x}_y$. There are three cases, depending on where $\bar{x}_{y'}$ falls.

- For $x \leq x' \leq \bar{x}_{y'} \leq \bar{x}_y$, relation (37) obviously holds with equality.
- For $x \leq \bar{x}_{y'} \leq x' \leq \bar{x}_y$, relation (37) becomes $f(x',y') + f(\bar{x}_y, y) \leq f(\bar{x}_{y'}, y') + f(\bar{x}_y, y)$ which is obvious by unconstrained maximality of $\bar{x}_{y'}$ for $f(\bar{x}_{y'}, y')$.
- For $\bar{x}_{y'} \leq x \leq x' \leq \bar{x}_y$, relation (37) becomes $f(x',y') + f(\bar{x}_y, y) \leq f(x,y') + f(\bar{x}_y, y)$. Indeed $f(x',y') \leq f(x,y')$ because $x' \geq x \geq \bar{x}_{y'}$, which is the unconstrained maximizer of the concave function $f(\cdot, y')$.

□

D.2. Irreversible allocation results and proofs

The next result extends Lemma 2 under irreversible commitment. In addition, it shows that revenue to go decreases with the number of committed make-goods, at a decreasing marginal rate. Moreover, there is a diminishing marginal rate of substitution between committed allocation and remaining target. The opportunity cost of a committed make-good is higher for lower performance targets (or the better the achieved performance). Finally, the marginal value of an additional make-good increases over time.

LEMMA 6. *The value function has the following properties:(a) J_t is decreasing in N_t and in x_t; (b) J_t is jointly concave in (x_t, N_t); (c) J_t has increasing differences in (x_t, N_t) and V_t has increasing differences in (x_{t+1}, N_t); (d) $V_t(x, N)$ and $J_t(x, N)$ have increasing differences in (x, t).*

Proof: (a) Monotonicity is easily proved by induction and Lemma 5 (a). The base case is trivial, transitions are linear and the profit per stage is state-independent.

(b) We show by backward induction that J_t is jointly concave in (x_t, N_t). The base case is trivial. Suppose J_{t+1} is jointly concave, so $J_{t+1}(x_{t+1}, N_t - \xi x_{t+1})$ is jointly concave in (x_{t+1}, N_t), hence so is $V_t(x_{t+1}, N_t)$. By Lemma 5 (b), J_t is jointly concave in (x_t, N_t).

(c) We prove both statements in parallel by induction. The base case is trivial. Assume for $t+1$ and show for t. Because J_{t+1} has increasing differences, by Lemma 4 (d), we obtain that $\mathbb{E}J_{t+1}(x_{t+1}, N_t - \xi_{t+1}x_{t+1})$ has increasing differences in (x_{t+1}, N_t), hence the same holds for V_t. Finally, increasing differences and concavity of V_t implies increasing differences of J_t by Lemma 5 (c).

(d) follows from the proof of Proposition 7 below. □

Extension of Proposition 5: The aggregation result follows the same lines as the proof of Proposition 5. The only additional condition to verify is feasibility of the multi-client solution provided by the equal GRP rule. We need to show that the lower bounds imposed by the irreversible allocation are satisfied. In a balanced GRP solution, $\frac{X_i^t}{N_i^t} = \frac{X_j^t}{N_j^t}$ for $N_i^t, N_j^t > 0$, so $\frac{X_i^t}{N_i^t - zX_i^t} = \frac{X_j^t}{N_j^t - zX_j^t}$, for any realization $z \geq 0$ of ξ, implying $\frac{X_i^t}{X_j^t} = \frac{N_i^{t+1}}{N_j^{t+1}} = \frac{X_i^{t+1}}{X_j^{t+1}}$. This shows that allocation increases proportionally over time for each unsatisfied client, so the irreversibility constraints are met. Hence the multi-client problem reduces to the aggregate case under irreversible commitment. □

Extension of Proposition 6: Monotonicity of the optimal policy in N and x follows from Lemma 6 (c) together with Topkis' Lemma. It remains to show monotonicity with time. We show by backwards induction the following three statements in parallel:

[J-t] $J_t(x, N - xz) - J_{t+1}(x, N - xz)$ decreasing in x for all $N, z \geq 0$.

[V-t] $V_{t-1}(x, N) - V_t(x, N)$ is decreasing in x for all $N \geq 0$.

[x-t] $x_t^*(x, N) \leq x_{t+1}^*(x, N)$ for all $N, x \geq 0$.

The base case [J-(T-1)] is obviously true. The following is true for all t:

[J-t] ⇒ [V-t] because $V_{t-1}(x, N) - V_t(x, N) = \mathbb{E}[J_t(x, N - x\xi) - J_{t+1}(x, N - x\xi)]$.

[V-t] ⇒ [x-t] by Topkis' Lemma.

It remains to show [x-t] ⇒ [J-(t-1)], i.e. the following function is decreasing in x for all $N, z \geq 0$:

$$\begin{aligned} j(x) &= J_{t-1}(x, N - xz) - J_t(x, N - xz) \\ &= V_{t-1}(x_t^*(x, N - xz), N - xz) - J_t(x, N - xz) \qquad (38) \\ &= \max_{x \leq y \leq Q} \pi(y) + \mathbb{E}[J_t(y, N - xz - y\xi) - J_t(x, N - xz)]. \qquad (39) \end{aligned}$$

By Proposition 7 we have

$$J_t(x, N-xz) = \begin{cases} J_t(0, N-xz), & \text{if } x^*_{t+1}(x, N-xz) > x \ (case\, 1) \\ V_t(x, N-xz), & \text{if } x^*_{t+1}(x, N-xz) = x \ (case\, 2). \end{cases} \tag{40}$$

Case 1: To show that $j(x)$ is decreasing in x, it is enough to show that the following function (inside the expectation in (39)) is decreasing in x for all $y, z, v \geq 0$:

$$\begin{aligned} g(x) &= J_t(y, N-xz-yv) - J_t(x, N-xz) = J_t(y, N-xz-yv) - J_t(0, N-xz) \\ &= [J_t(y, N-xz-yv) - J_t(y, N-xz)] + [J_t(y, N-xz) - J_t(0, N-xz)]. \end{aligned} \tag{41}$$

Indeed, by Lemma 6, the first difference is decreasing in x by concavity of J_t in the second argument, and the second difference is decreasing in x by increasing differences of J_t.

Case 2: We have $x \leq x^*_t(x, N-xz) \leq x^*_{t+1}(x, N-xz) = x$ by the induction hypothesis [x-t]. Therefore $x^*_t(x, N-xz) = x$. This together with (40) allows to rewrite (38) as

$$j(x) = V_{t-1}(x, N-xz) - V_t(x, N-xz) = \mathbb{E}[J_t(x, N-xz-x\xi) - J_{t+1}(x, N-xz-x\xi)].$$

The right hand side is decreasing in x because the function inside the expectation $J_t(x, N-x(z+v)) - J_{t+1}(x, N-x(z+v))$ is so for any $z, v \geq 0$, by [J-t]. This concludes the proof. □

Appendix E: Proof of Proposition 8 (Service models)

Part (a) is obvious, so the proof focuses on part (b). Once the firm accepted a (feasible) set of contracts I with performance targets N_i, the multi-client inventory planning problem amounts to setting $x^*_i = N_i/F^{-1}(\epsilon) = q_i$, the minimum inventory that satisfies the service constraint. Similar to the result of Proposition 1, the resulting total allocation to the upfront market $x^* = \sum_{i\in I} x^*_i$ can be obtained by solving a service model (S) with one aggregate service constraint corresponding to the cumulative target $N = \sum_{i\in I} N_i$. The resulting upfront allocation $x^* = N/F^{-1}(\epsilon)$ is then divided among clients in proportion to their contracted performance targets N_i, i.e. $N_i/x^*_i = N/x^* = F^{-1}(\epsilon)$. Like for the penalty model, the firm only has to set GRP allocation (in this case $F^{-1}(\epsilon)$) for the entire upfront market, and equate this across clients.

If $y^*_i = 0$ in an optimal solution of model (24), then it is optimal to set $x^*_i = 0$. Otherwise, the ith service constraint needs to hold, moreover it is binding, i.e. $x^*_i = N_i y^*_i / F^{-1}(\epsilon) = q_i y^*_i$. Model (24) can be simplified as follows:

$$\begin{gathered} \max\ pQ + \textstyle\sum_i (B_i - pq_i) y_i \\ \textstyle\sum q_i y_i \leq Q, y_i \in Y \end{gathered} \tag{42}$$

This is a knapsack problem, whose fractional solution serves clients in decreasing order of $\dfrac{B_i - q_i}{q_i} \sim \dfrac{B_i}{q_i} \sim \dfrac{B_i}{N_i} = C_i$. Clients are profitable as long as $p \leq B_i/q_i = C_i F^{-1}(\epsilon)$. □

Europe Campus
Boulevard de Constance,
77305 Fontainebleau Cedex, France
Tel: +33 (0)1 6072 40 00
Fax: +33 (0)1 60 74 00/01

Asia Campus
1 Ayer Rajah Avenue, Singapore 138676
Tel: +65 67 99 53 88
Fax: +65 67 99 53 99

www.insead.edu

INSEAD

The Business School
for the World