Choice network revenue management based on new compact formulations

Sumit Kunnumkal* Kalyan Talluri†

August 19, 2015

Abstract

The choice network revenue management model incorporates customer purchase probabilities as a function of the offered products, and is the appropriate model for airline and hotel network revenue management, dynamic sales of bundles, and dynamic assortment optimization. The optimization problem is a stochastic dynamic program and is intractable; a linear programming approximation called choice deterministic linear program (CDLP) is usually used to generate controls. Tighter approximations such as affine and piecewise-linear relaxations have been proposed, but their complexity for the simplest choice model, the multinomial logit (MNL) model with a single segment, was unknown. We first show that the affine relaxation (and hence the piecewise-relaxation) is NP-hard even for a single-segment MNL choice model. By analyzing the affine relaxation we derive a new linear programming approximation that admits a compact representation, implying tractability. We prove that its value falls between the CDLP value and the affine relaxation value, often coming close to the latter in our numerical experiments. This is the first tractable relaxation for the choice network revenue management problem that is provably tighter than CDLP. This formulation in turn leads to new policies that, in our numerical experiments, show promise of significant increases in revenue, 2% on average over CDLP. Our analysis also yields a more complicated tractable relaxation, as well as a hierarchical family of relaxations that approach the affine relaxation. We give extensions to the case with multiple customer segments with overlapping consideration sets where choice by each segment is according to the MNL model.

*Indian School of Business, Hyderabad, 500032, India, email: sumit.kunnumkal@isb.edu
†Imperial College Business School, South Kensington Campus, London SW7 2AZ, U.K. email: kalyan.talluri@imperial.ac.uk
1 Introduction and literature review

In this paper we consider Network Revenue Management (NRM) under a choice model of consumer behavior—a problem relevant for the airline, advertising, hotel and car rental industries—and address its optimization.

Revenue Management (RM) controls the sale of different products to maximize revenue, and in NRM, products consume multiple resources creating network dependencies. In the canonical airline example, resources correspond to flight legs and products correspond to itineraries that span multiple flight legs; for hotels, resources correspond to hotel rooms for each night and products correspond to multi-night stays; in advertising, resources correspond to ad slots and products correspond to bundles of ad slots. The network dependencies introduce a considerable amount of additional complexity in the stochastic control problem.

The NRM problem can be formulated as a stochastic dynamic program. However, solving the Bellman equation is intractable even for very small problems because of an explosion of the state space. Considering the intractability of the NRM dynamic program, Gallego, Iyengar, Phillips, and Dubey [4] and Liu and van Ryzin [7] proposed a linear programming approximation called the choice deterministic linear program (CDLP) (similar to some earlier deterministic approximations proposed for solving NRM under the simpler perfect segmentation assumption; see Talluri and van Ryzin [15]). The optimal objective function value of CDLP gives an upper bound on the value function of the NRM dynamic program. Upper bounds are useful both for deriving controls from them, as well as for assessing the sub-optimality of policies.

The CDLP, however, has a drawback—the number of columns are exponential in the number of products; so it has to be solved using column generation. Liu and van Ryzin [7] show that the CDLP column generation procedure is tractable for the multinomial logit (MNL) choice model with multiple customer segments when the customers’ consideration sets do not overlap. More recently, Gallego, Ratliff, and Shebalov [5] show that CDLP has a compact linear programming formulation under the MNL model with disjoint consideration sets. On the other hand, for the problem with two segments whose consideration sets overlap, CDLP is intractable even for the MNL model (Bront, Méndez-Díaz, and Vulcano [3], Rusmevichientong, Shmoys, Tong, and Topaloglu [12]). Zhang and Adelman [17] investigate an affine relaxation to the NRM dynamic program and show that it obtains a tighter upper bound than CDLP.

This paper builds on these advances and makes the following research contributions:
1. We show that the affine relaxation of NRM is NP-hard even for the single-segment MNL model (perhaps the simplest of choice models). Our result implies that stronger solution methods that obtain tighter bounds than the affine relaxation (such as the piecewise-linear approximation proposed by Meissner and Strauss [9]) are also NP-hard for the single-segment MNL model. On the other hand, our hardness result motivates solution methods that tighten the CDLP bound and remain tractable, at least for the single-segment MNL model.

2. We propose new, compact, linear programming formulations that give a tighter bound on the dynamic program value function than CDLP, improving upon the work of Gallego et al. [5]. Compact formulations are attractive from an implementation perspective since they do not require customized coding for constraint-separation or column-generation. To our knowledge, these are the first tractable approximation methods for MNL that are provably tighter than CDLP. In numerical experiments they typically produce upper bounds that are close to the affine bound (reaching nearly 75% of the affine bound on average) and have good revenue performance (obtaining on average above 95% of the revenue possible from policies based on the affine relaxation).

3. We show how our ideas can be extended to the mixture-of-multinomial-logits (MMNL) model (McFadden and Train [8]), with disjoint as well as with overlapping consideration sets.

4. We propose control policies based on the new formulations and test their performance through an extensive numerical study. Our methods yield noticeable benefits both in terms of tighter bounds (over 1.5% above CDLP on average across instances) and improved revenue performance (over 2% above CDLP on average across instances). The benefits primarily come from sharper value function approximations towards the end of the selling horizon when capacity tends to be relatively scarce. So one option for practitioners is to switch to these methods during the last few days of the sales horizon.

The remainder of the paper is organized as follows: In §2 we describe the choice NRM model, the notation, and the basic dynamic program, the CDLP and the affine relaxation of the NRM dynamic program. Next, in §3 we show that the affine relaxation is NP-hard even for the single-segment MNL model. We describe our first tractable approximation method in §4 and build on it to obtain tractable, tighter approximations in §5. §6 discusses extensions to the MMNL model. §7 contains our computational study using the new formulations.
2 Problem formulation

We are interested in controlling the sale of products over a finite sales horizon. A product is a specification of a price and the set of resources that it consumes. Time is discrete and the sales horizon consists of \( \tau \) intervals, indexed by \( t \). The sales horizon begins at time \( t = 1 \) and ends at \( t = \tau \); all the resources perish instantaneously at time \( \tau + 1 \). We make the standard assumption that the time intervals are fine enough so that the probability of more than one customer arriving in any single time period is negligible.

We let \( \mathcal{I} \) denote the set of resources and \( \mathcal{J} \) the set of products. We index resources by \( i \) and products by \( j \). We let \( f_j \) denote the revenue associated with product \( j \) and use \( \mathcal{I}_j \subseteq \mathcal{I} \) to denote the set of resources used by product \( j \). We let \( 1_{[\cdot]} \) denote the indicator function, 1 if true and 0 if false and \( 1_{[\mathcal{I}_j]} \) denote the vector of resources used by product \( j \), with a 1 in the \( i \)th position if \( i \in \mathcal{I}_j \) and a 0 otherwise. We use \( \mathcal{J}_i \subseteq \mathcal{J} \) to denote the set of products that use resource \( i \).

In each period the firm offers a subset \( S \) of its products for sale, called the offer set. We write \( i \in \mathcal{I}_S \) whenever there is a \( j \in S \) with \( i \in \mathcal{I}_j \); that is, there is at least one product in the offer set \( S \) that uses resource \( i \).

We use superscripts on vectors to index the vectors (for example, the resource capacity vector associated with time period \( t \) would be \( r^t \)) and subscripts to indicate components (for example, the capacity on resource \( i \) in time period \( t \) would be \( r^t_i \)). Therefore, \( r^1 = [r^1_i] \) represents the initial capacity on the resources and \( r^t = [r^t_i] \) denotes the remaining capacity on the resources at the beginning of time period \( t \). The remaining capacity \( r^t_i \) takes values in the set \( \mathcal{R}_i = \{0, \ldots, r^1_i\} \) and \( \mathcal{R} = \prod \mathcal{R}_i \) represents the state space at each time \( t \).

2.1 Demand model

We have multiple customer segments, each with distinct purchase behavior. We let \( \mathcal{L} \) denote the set of customer segments. In each period a customer from segment \( l \in \mathcal{L} \) arrives with probability \( \lambda_l \) so that \( \lambda = \sum_l \lambda_l \) is the total arrival rate. Note that conditioned on a customer arrival, \( \lambda_l/\lambda \) is the probability that the customer belongs to segment \( l \).

Customer segment \( l \) has a consideration set \( \mathcal{C}_l \subseteq \mathcal{J} \) of products that it considers for purchase. We assume this consideration set is known to the firm (by a previous process of estimation and
The choice probabilities of a segment-$l$ customer are not affected by products not in its consideration set. Given an offer set $S$, an arriving customer in segment $l$ purchases a product $j$ in the set $S_l = C_l \cap S$ or leaves without making a purchase. The no-purchase option is indexed by 0 and is always present for the customer.

Within each segment, choice is according to the MNL model. The MNL model associates a preference weight with each alternative including the no-purchase alternative. We let $w_{lj}$ denote the preference weight associated with a segment-$l$ customer for product $j$. Without loss of generality, by suitably normalizing the weights, we set the no-purchase weight $w_{l0}$ to be 1. The probability that a segment-$l$ customer purchases product $j$ when $S$ is the offer set is

$$P_{lj}(S) = \frac{w_{lj}[j \in S_l]}{1 + \sum_{k \in S_l} w_{lk}}. \quad (1)$$

The probability that the customer does not purchase anything is $P_{l0}(S) = 1/(1 + \sum_{k \in S_l} w_{lk})$. We note that the preference weights are inputs to our model; estimating them is outside the scope of the paper. We refer the reader to Ben-Akiva and Lerman [2] for further background on this popular choice model.

Given a customer arrival, and an offer set $S$, the probability that the firm sells $j \in S$ is given by $P_j(S) = \sum_l \lambda_l P_{lj}(S)$ and makes no sale with probability $P_b(S) = 1 - \sum_{j \in S} P_j(S)$. The expected sales for product $j$ is therefore $\lambda P_j(S) = \sum_l \lambda_l P_{lj}(S)$, while $1 - \lambda + \lambda P_b(S) = 1 - \sum_{j \in S} \lambda P_j(S)$ is the probability of no sales in a time period. Given an offer set $S$, $Q_i(S) = \sum_{j \in J_i} P_{lj}(S)$ denotes the expected capacity consumed on resource $i$ conditional on a segment-$l$ customer arrival and $Q_i(S) = \sum_l \frac{1}{\lambda_l} Q_{li}(S)$ denotes the expected capacity consumed on resource $i$ conditional on a customer arrival. Note that $\lambda Q_i(S) = \sum_l \lambda_l Q_{li}(S)$ gives the expected capacity consumed on resource $i$ in a time period. The revenue functions can be written as $R_i(S) = \sum_{j \in S} f_j P_{lj}(S)$ and $R(S) = \sum_{j \in S} f_j P_j(S)$.

We assume that the arrival rates and choice probabilities are stationary. This is for brevity of notation only, all our results go through with non-stationary arrival rates and choice probabilities.

### 2.2 Choice dynamic program

The dynamic program (DP) to determine optimal controls is as follows. Let $V_t(r^t)$ denote the maximum expected revenue to go, given remaining capacity $r^t$ at the beginning of period $t$. Then
$V_t(r')$ must satisfy the Bellman equation

$$V_t(r') = \max_{S \subseteq S(r')} \left\{ \sum_{j \in S} \lambda P_j(S) \left[ f_j + V_{t+1}(r' - 1_{[\mathbf{z}_{ij}]} \right] + \left[ \lambda P_0(S) + 1 - \lambda \right] V_{t+1}(r') \right\},$$

(2)

where

$$S(r) = \{ j \mid 1_{[\mathbf{z}_{ij}]} \leq r_i \forall i \}$$

represents the set of products that can be offered given the capacity vector $r$. The boundary conditions are $V_{r+1}(r) = V_t(0) = 0$ for all $r$ and for all $t$, where 0 is a vector of all zeroes. $V^{DP} = V_t(r^1)$ denotes the optimal expected total revenue over the sales horizon, given the initial capacity vector $r^1$.

### 2.3 Linear programming formulation of the dynamic program

The value functions can, alternatively, be obtained by solving a linear program (LP). The LP formulation of (2) has a decision variable for each state vector in each period $V_t(r)$ and is as follows:

$$V^{DPLP} = \min_{V} V_1(r^1)$$

(DPLP) s.t. $V_t(r) \geq \sum_{j} \lambda P_j(S) \left[ f_j + V_{t+1}(r - 1_{[\mathbf{z}_{ij}]} - V_{t+1}(r) \right] + V_{t+1}(r)$

$$\forall r \in \mathcal{R}, S \subseteq S(r), t.$$ 

Both dynamic program (2) and DPLP are computationally intractable, but DPLP turns out to be useful in developing value function approximation methods, as shown in Zhang and Adelman [17].

### 2.4 Choice deterministic LP

The choice deterministic linear program (CDLP) proposed in Gallego et al. [4] and Liu and van Ryzin [7] is a certainty-equivalence approximation to (2). We write CDLP as the following LP:

$$V^{CDLP} = \max_{h} \sum_{i} \sum_{S} \lambda R(S) h_{S,t}$$

(CDLP) s.t. $\sum_{k=1}^{t} \sum_{S} \lambda Q_i(S) h_{S,k} \leq r^1_i \forall i, t$ \hspace{1cm} (4)

$$\sum_{S} h_{S,t} = 1 \forall t$$ \hspace{1cm} (5)

$$h_{S,t} \geq 0 \forall S, t.$$ 

6
The decision variable \( h_{S,t} \) can be interpreted as the frequency with which set \( S \) (including the empty set) is offered at time period \( t \). The first set of constraints ensure that the total expected capacity consumed on resource \( i \) up until time period \( t \) does not exceed the available capacity. Note that since \( h_{S,t} \geq 0 \), constraints (4) are redundant except for the last time period. Still, this expanded formulation is useful when we compare CDLP with other approximation methods. The second set of constraints states that the sum of the frequencies adds up to 1.

The dual of CDLP turns out to be useful in our analysis. Associating dual variables \( \gamma = \{\gamma_{i,t} | \forall i, t\} \) with constraints (4) and \( \beta = \{\beta_t | \forall t\} \) with constraints (5), the dual of CDLP is

\[
V^{dCDLP} = \min_{\beta, \gamma} \sum_t \beta_t + \sum_t \sum_i \gamma_{i,t} r_i
\]

\[(dCDLP) \text{ s.t.} \quad \beta_t + \sum_i \left( \sum_{k=t}^T \gamma_{i,k} \right) \lambda Q_i(S) \geq \lambda R(S) \quad \forall t, S \]

\( \gamma_{i,t} \geq 0 \quad \forall i, t. \) (6)

Liu and van Ryzin [7] show that the optimal objective function value of CDLP, \( V^{CDLP} \), is an upper bound on \( V^{DPLP} \). Besides giving an upper bound on the value function, CDLP can also be used to construct different heuristic control policies. Let \( \bar{\gamma} = \{\bar{\gamma}_{i,t} | \forall i, t\} \) denote the optimal values of the dual variables associated with constraints (4), we interpret \( \bar{\gamma}_{i,t} \) as giving the value of an additional unit of capacity on resource \( i \) from time period \( t \) to \( t + 1 \). With this interpretation, \( \sum_{s=t}^T \bar{\gamma}_{i,s} \) gives the marginal value of capacity on resource \( i \) at time period \( t \). Zhang and Adelman [17] approximate the value function as

\[
\hat{V}_t(r) = \sum_i \left( \sum_{s=t}^T \bar{\gamma}_{i,s} \right) r_i \quad (7)
\]

and if \( r^t \) is the vector of remaining resource capacities at time \( t \), solve the problem

\[
\max_{S \subseteq S(r^t)} \left\{ \sum_{j \in S} \lambda P_j(S) \left[ f_j + \hat{V}_{t+1} \left( r^t - \textbf{1}[j] \right) \right] + \lambda P_0(S) + 1 - \lambda \hat{V}_{t+1} \left( r^t \right) \right\}, \quad (8)
\]

and use the policy of offering the set that achieves the maximum in the above optimization problem.

The number of decision variables in CDLP is exponential in the number of products and so it has to be solved using column generation. The tractability of column generation depends on the underlying choice model. Liu and van Ryzin [7] show that the column generation procedure can be efficiently carried out when choice is according to the MNL model and the consideration sets of the different segments do not overlap. That is, we have \( C_l \cap C_m = \emptyset \) for segments \( l \) and \( m \). Under the
same set of assumptions, Gallego et al. [5] further show that CDLP has the following equivalent, compact formulation

\[
V^{SBLP} = \max_t \sum_t \sum_{j \in \mathcal{C}_t} \lambda_t f_j x_{j,t} \\
(SBLP) \text{ s.t.} \sum_t \sum_{j \in \mathcal{J}_t \cap \mathcal{C}_t} \lambda_t x_{j,t} \leq r_{i,t}^l \quad \forall i, t
\]

\[
x_{0,t} + \sum_{j \in \mathcal{C}_t} x_{j,t} = 1 \quad \forall i, t
\]

\[
\frac{x_{j,t}}{w_{j,t}} - x_{0,t} \leq 0 \quad \forall l, j \in \mathcal{C}_t, t
\]

\[
x_{0,t}, x_{j,t} \geq 0 \quad \forall l, j, t.
\]

In the above sales-based linear program (SBLP), the decision variables \(x_{j,t}\) can be interpreted as the sales rate for product \(j\) at time \(t\). Note that SBLP is a compact formulation since the number of constraints and decision variables is polynomial in the number of products and resources. On the other hand, if the consideration sets overlap, Bront et al. [3] and Rusmevichientong et al. [12] show that the CDLP column generation is NP-hard even under the MNL choice model.

2.5 Affine relaxation

The second approximation method we consider is the affine relaxation, where the value function is approximated as \(V_t(r) = \theta_t + \sum_i V_{i,t} r_i\). Note that \(V_{i,t}\) can be interpreted as the marginal value of capacity on resource \(i\) at time \(t\). Substituting this value function approximation into the formulation DPLP we get the affine relaxation LP

\[
V^{AF} = \min_{\theta, V} \theta_t + \sum_i V_{i,t} r_i^l
\]

\[
(AF) \text{ s.t.} \quad \theta_t + \sum_i V_{i,t} r_i \geq \sum_j \lambda P_j(S) \left[ f_j - \sum_{i \in \mathcal{I}_j} V_{i,t+1} \right] + \theta_{t+1} + \sum_i V_{i,t+1} r_i \\
\forall r \in \mathcal{R}, S \subseteq S(r), t
\]

\[
\theta_t \geq 0, V_{i,t} \geq 0 \quad \forall i, t
\]

with the boundary conditions \(\theta_{t+1} = 0, V_{i,t+1} = 0\). Zhang and Adelman [17] show that the optimal objective function value \(V^{AF}\) is an upper bound on the value function and that there exists an optimal solution \((\hat{\theta}, \hat{V})\) of AF that satisfies \(\hat{V}_{i,t} - \hat{V}_{i,t+1} \geq 0\) for all \(i\) and \(t\).

While the number of decision variables in AF is manageable, the number of constraints is ex-
ponential both in the number of products as well as the number of resources. Vossen and Zhang [16] use Dantzig-Wolfe decomposition to derive a reduced, equivalent formulation of \( AF \), where the number of constraints is exponential only in the number of products.

We give an alternative, simpler proof of the reduction below. The analysis we present also turns out to be useful in the development of our tractable solution methods later. We make a change of variables \( t_i = t_i + 1 \), and \( \gamma_{i,t} = v_{i,t} - v_{i,t+1} \) and write \( AF \) equivalently as

\[
\min_{\beta, \gamma} \sum_t \beta_t + \sum_t \sum_i \gamma_{i,t} r_i^1 \\
\text{s.t.} \quad \beta_t + \sum_i \gamma_{i,t} r_i + \sum_j \lambda p_j(S) \left( \sum_{i \in I_j} \sum_{k=t+1}^\tau \gamma_{i,k} \right) - f_j \geq 0 \quad \forall r \in R, S \subseteq S(r), t
\]

(10)

where we use the fact that \( v_{i,t} = \sum_{k=t}^{\tau} \gamma_{i,k} \) and so \( \sum_{k=t}^{\tau} \gamma_{i,k} \) can be interpreted as the marginal value of capacity on resource \( i \) at time \( t \). Note that the nonnegativity constraint on \( \gamma_{i,t} \) is without loss of generality, since there exists an optimal solution to \( AF \) that satisfies \( v_{i,t} - v_{i,t+1} \geq 0 \).

Now, constraints (10) can be written as

\[
\min_{r \in R, S \subseteq S(r)} \left\{ \beta_t + \sum_i \gamma_{i,t} r_i + \sum_j \lambda p_j(S) \left( \sum_{i \in I_j} \sum_{k=t+1}^\tau \gamma_{i,k} \right) - f_j \right\} \geq 0
\]

(11)

for all \( t \). Since \( \gamma_{i,t} \geq 0 \), the coefficient of \( r_i \) in minimization problem (11) is nonnegative, and we can assume \( r_i \in \{0, 1\} \) in the minimization (as larger values of \( r_i \) would be redundant in \( S \subseteq S(r) \) and would only increase the objective value). Moreover, since \( \gamma_{i,t} \geq 0 \), for any set \( S \), we have \( r_i = 0 \) for \( i \notin I_S \). On the other hand, feasibility requires we have \( r_i = 1 \) for \( i \in I_S \). Therefore, (11) can be written as

\[
\min_S \left\{ \beta_t + \sum_i \mathbb{1}_{i \in I_S} \gamma_{i,t} + \sum_j \lambda p_j(S) \left( \sum_{i \in I_j} \sum_{k=t+1}^\tau \gamma_{i,k} \right) - f_j \right\} \geq 0
\]

and we can write \( AF \) equivalently as

\[
V^{RAF} = \min_{\beta, \gamma} \sum_t \beta_t + \sum_t \sum_i \gamma_{i,t} r_i^1 \\
(\text{RAF}) \text{s.t.} \quad \beta_t + \sum_i \mathbb{1}_{i \in I_S} \gamma_{i,t} + \sum_j \left( \sum_{k=t+1}^\tau \gamma_{i,k} \right) \lambda q_i(S) \geq \lambda R(S) \quad \forall t, S
\]

(12)

Notice that the number of constraints in the reduced formulation \( RAF \) is an order of magnitude smaller than \( AF \). Taking the dual of \( RA \) by associating dual variables \( h_{S,t} \) with constraints (12),
we get
\[ V^{\text{dRAF}} = \max_h \sum_t \sum_S \lambda R(S) h_{S,t} \]
\[ (\text{dRAF}) \text{ s.t.} \]
\[ \sum_S \left( \sum_{k=1}^{t-1} \lambda Q_i(S) h_{S,k} + \mathbb{1}_{[i \in I_S]} h_{S,t} \right) \leq r_i^t \quad \forall i, t \]
\[ \sum_S h_{S,t} = 1 \quad \forall t \]
\[ h_{S,t} \geq 0 \quad \forall S, t. \]

The above arguments imply that

**Proposition 1.** *(Vossen and Zhang [16])* \( V^{AF} = V^{RAF} = V^{dRAF} \).

We close this section with two remarks. First, in addition to giving an upper bound on the optimal expected total revenue, the affine relaxation can also be used to construct heuristic control policies. Letting \( (\hat{\beta}, \hat{\gamma}) \), with \( \hat{\beta} = \{ \hat{\beta}_i \mid \forall t \} \) and \( \hat{\gamma} = \{ \hat{\gamma}_{i,t} \mid \forall i, t \} \), denote an optimal solution to RAF, we use \( \sum_{i=1}^T \hat{\gamma}_{i,k} \) to approximate the marginal value of capacity on resource \( i \) at time \( t \). We approximate \( V_i(r) \) using (7) and solve problem (8) using this value function approximation to decide on the set of products to be offered at time period \( t \). Second, Zhang and Adelman [17] show that the upper bound obtained by AF is tighter than CDLP. In that sense, AF is a better approximation than CDLP. At the same time, it is important to understand the computational effort required by AF to obtain a tighter bound. We explore this question in the following section.

### 3 Tractability of the affine relaxation for MNL with a single segment

In this section, we focus on the tractability of the affine relaxation for the single-segment MNL model. We restrict our attention to the single-segment MNL since it is one of the few cases where CDLP is tractable. We show that the affine relaxation is NP-hard even for this simple choice model.

Let the preference weights be \( w_j \) (as mentioned earlier, we drop the segment index \( l \) when we are analyzing a single segment problem). The choice probabilities, expected resource consumptions and expected revenues are then given by

\[ P_j(S) = \frac{\mathbb{1}_{[j \in S]} w_j}{1 + \sum_{k \in S} w_k} \]
\[ Q_i(S) = \frac{\sum_{j \in I_i \cap S} w_j}{1 + \sum_{j} w_j} \]
\[ R(S) = \frac{\sum_{j} f_j w_j}{1 + \sum_{j} w_j}. \]

(13)
Since RAF has an exponential number of constraints, we have to use constraint separation, and generate constraints (12) violated by a solution on the fly. Following the result of Grötschel, Lovász, and Schrijver [6], polynomial-solvability of an LP is equivalent to polynomial-time generation of violated constraints and so we focus on separating constraints (12).

Substituting (13) into constraint (12), we obtain
\[
\beta_t + \gamma_{S,t} + \sum_i \left( \sum_{k=i+1}^r \gamma_{i,k} \right) \lambda \frac{\sum_{j \in S \cap I_i w_j}{1 + \sum_{j \in S} w_j}}{1 + \sum_{j \in S} w_j} \geq \lambda \frac{\sum_{j \in S} f_j w_j}{1 + \sum_{j \in S} w_j}
\]
where
\[
\gamma_{S,t} = \sum_i \mathbb{1}_{i \in I_S} \gamma_{i,t}.
\]
Multiplying both sides by the positive quantity \(1 + \sum_{j \in S} w_j\) and simplifying, constraint (12) of RAF can be equivalently written as
\[
\beta_t \geq -\gamma_{S,t} \left(1 + \sum_{j \in S} w_j\right) - \sum_{j \in S} \zeta_{j,t}(\beta, \gamma),
\]
where
\[
\zeta_{j,t}(\beta, \gamma) = w_j \left[ \beta_t + \lambda \left( \sum_{i \in I_j} \sum_{k=i+1}^r \gamma_{i,k} - f_j \right) \right].
\]
Since the constraint has to be satisfied for every \(S\) and \(t\), we have \(\beta_t \geq \Pi_t^{AF}(\beta, \gamma)\) for all \(t\), where
\[
\Pi_t^{AF}(\beta, \gamma) = \max_S \left\{ -\gamma_{S,t} \left(1 + \sum_{j \in S} w_j\right) - \sum_{j \in S} \zeta_{j,t}(\beta, \gamma) \right\}
\]
and the affine relaxation constraint (12) can be equivalently written as
\[
\beta_t \geq \Pi_t^{AF}(\beta, \gamma) \quad \forall t.
\]

Generating constraints on the fly involves checking, given a set of values \((\beta, \gamma)\), if constraint (14) is satisfied for all \(S\). If not, we add the violated constraint to the LP. In other words, the RAF separation problem at time \(t\) involves solving optimization problem (15) and determining if \(\beta_t \geq \Pi_t^{AF}(\beta, \gamma)\). If \(\beta_t \geq \Pi_t^{AF}(\beta, \gamma)\), then constraint (14) is satisfied for all \(S\) at time \(t\). Otherwise, the set \(S^*\) which attains the maximum in problem (15) violates the constraint, and we add the constraint for set \(S^*\) to the LP.

Proposition 2 below states that the affine relaxation separation problem for MNL with a single segment, as given in (14), is NP-hard.
Proposition 2. The following problem is NP-complete:
Input: \( w_j \geq 0, \ 1 \geq \lambda \geq 0, \ f_j \geq 0, \) and values \( \beta_t \) and \( \gamma_{i,t} \geq 0. \)
Question: Is there a set \( S \) that violates (14)?

Proof
Proof given in the Online Appendix.

Therefore, even though the affine relaxation tightens the CDLP bound, it comes at a significant cost. This motivates the solution method that we propose in the following section, which tightens the CDLP bound while retaining tractability.

4 New improved relaxation

In this section we propose our first tractable approximation method that tightens the CDLP bound. We also show that our approximation method can, in fact, be formulated as a compact LP. In our initial development, we restrict attention to the single-segment MNL choice model. We emphasize that this is only for clarity of exposition. In §6 we show how the ideas can be readily extended to more realistic variants of the MNL model that consider multiple customer segments. Moreover, all of the test problems in our computational experiments involve multiple customer segments.

4.1 Preliminaries

All of our approximation methods involve solving an optimization problem of the form \( \min_{\beta, \gamma} \sum_t \beta_t + \sum_t \sum_i \gamma_{i,t} r_{i,t}^1 \) subject to the constraints \( \beta_t \geq \Pi_t(\beta, \gamma), \) where \( \Pi_t(\cdot, \cdot) \) is a scalar function of \( \beta = \{\beta_t | \forall t\} \) and \( \gamma = \{\gamma_{i,t} | \forall i, t\}. \) The following observation is useful in comparing the upper bounds obtained by the different approximation methods.

Lemma 1. Let
\[
V^I = \min_{\beta, \gamma} \sum_t \beta_t + \sum_t \sum_i \gamma_{i,t} r_{i,t}^1
\]
(I) s.t. \( \beta_t \geq \Pi_t^I(\beta, \gamma), \ \gamma_{i,t} \geq 0 \ \forall i, t, \)
and

\[ V^{II} = \min_{\beta, \gamma} \sum_{t} \beta_t + \sum_{t} \sum_{i} \gamma_{i,t} r_{i,t}^{1} \]

\[ (II) \quad \text{s.t.} \quad \beta_t \geq \Pi^{II}_t(\beta, \gamma), \quad \gamma_{i,t} \geq 0 \quad \forall i, t. \]

If \( \Pi^{II}_t(\beta, \gamma) \leq \Pi^{II}_t(\beta, \gamma) \) for all \( t \), then \( V^I \leq V^{II} \).

**Proof**

The proof follows by noting that a feasible solution to problem (II) is also feasible to problem (I) and both optimization problems have the same objective function.

\[ \square \]

### 4.2 CDLP vs. AF for single-segment MNL

We begin by comparing the CDLP and AF separation problems for the single-segment MNL model. For this choice model, the CDLP constraints can be separated efficiently, while the AF separation problem is intractable. Comparing the CDLP and AF separation problems helps us identify the difficult term in the affine relaxation. Replacing this difficult term in the AF separation problem with a more tractable term yields our approximation method.

Using the single-segment MNL formulas for the expected resource consumptions and expected revenues, the CDLP dual constraint (6) can be written as

\[ \beta_t \geq -\sum_{j \in S} w_j \left( \beta_t + \lambda \left( \sum_{i} \sum_{k=t}^{\tau} \gamma_{i,k} - f_j \right) \right) \quad \forall t, S \]

which looks similar to the right-hand-side of (14) except that the inner summation over \( k \) runs from \( t \) instead of \( t + 1 \). To make the comparison with AF easier, we rewrite the above constraint as

\[ \beta_t \geq \Pi^{CDLP}_t(\beta, \gamma) \quad \forall t \]

where

\[ \Pi^{CDLP}_t(\beta, \gamma) = \max_{S} \left\{ -\lambda \sum_{j \in S} w_j \left( \sum_{i \in I_j} \gamma_{i,t} \right) - \sum_{j \in S} \zeta_{j,t}(\beta, \gamma) \right\}. \]

Since \( 0 \leq \lambda \leq 1 \), and \( \gamma_{S,t} = \sum_{j \in I_S} \gamma_{i,t} \geq \sum_{i \in I_j} \gamma_{i,t} \geq 0 \) for all \( j \in S \), we have

\[ \gamma_{S,t} \left( 1 + \sum_{j \in S} w_j \right) \geq \lambda \sum_{j \in S} w_j \left( \sum_{i \in I_j} \gamma_{i,t} \right). \]
Therefore $\Pi_t^{AF}(\beta, \gamma) \leq \Pi_t^{CDLP}(\beta, \gamma)$ and by Lemma 1, $V_t^{AF} \leq V_t^{CDLP}$, which gives an alternative proof of the $AF$ bound being tighter than the $CDLP$ bound. More importantly, the comparison hints at how we can obtain tractable relaxations that are tighter than $CDLP$.

### 4.3 A tractable approximation

We are now ready to describe our first tractable approximation method, which we refer to as weak affine relaxation ($wAR$). The difficult term in (15) is the $\gamma_{st}(1 + \sum_{j \in S} w_j)$, and $CDLP$ is tractable as it replaces this by $\lambda \sum_{j \in S} w_j \left(\sum_{i \in I_j} \gamma_{i,t}\right)$. We instead replace the $\gamma_{st}(1 + \sum_{j \in S} w_j)$ term in (15) with $\gamma_{st} + \sum_{j \in S} w_j \left(\sum_{i \in I_j} \gamma_{i,t}\right)$ and solve the LP

$$V_t^{wAR} = \min_{\beta, \gamma} \left\{ \sum_t \beta_t + \sum_t \sum_i \gamma_{i,t} r_i^1 \right\} \quad (wAR) \text{ s.t.} \quad \beta_t \geq \Pi_t^{wAR}(\beta, \gamma) \quad \forall t$$

$$\gamma_{i,t} \geq 0 \quad \forall i, t,$$

where

$$\Pi_t^{wAR} = \max_S \left\{ -\gamma_{S,t} - \sum_{j \in S} w_j \left(\sum_{i \in I_j} \gamma_{i,t}\right) - \sum_{j \in S} \zeta_{j,t}(\beta, \gamma) \right\}. \quad (20)$$

Proposition 3 below shows that $wAR$ obtains an upper bound on the value function that is weaker than $AF$ but stronger than $CDLP$.

**Proposition 3.** $V_t^{AF} \leq V_t^{wAR} \leq V_t^{CDLP}$.

**Proof.**

The proof follows by noting that

$$\gamma_{S,t} \left(1 + \sum_{j \in S} w_j\right) \geq \gamma_{S,t} + \sum_{j \in S} w_j \left(\sum_{i \in I_j} \gamma_{i,t}\right) \geq \lambda \sum_{j \in S} w_j \left(\sum_{i \in I_j} \gamma_{i,t}\right).$$

Therefore $\Pi_t^{AF}(\beta, \gamma) \leq \Pi_t^{wAR}(\beta, \gamma) \leq \Pi_t^{CDLP}(\beta, \gamma)$ and the result now follows from Lemma 1.

In the remainder of this section, we show that the weak affine relaxation upper bound, $V_t^{wAR}$, can be obtained in a tractable manner; moreover we show that the weak affine relaxation LP can, in fact, be reformulated as a compact LP where the number of variables and constraints is polynomial in the number of products and resources.
Observe that solving problem (20) in an efficient manner is key to separating the weak affine relaxation constraints efficiently. Therefore, we focus on solving optimization problem (20). Introducing decision variables $q_{i,t}$ and $u_{j,t}$, respectively, to indicate if resource $i$ and product $j$ are open at time $t$, problem (20) can be formulated as the integer program

$$
\Pi_t^{wAR}(\beta, \gamma) = \max \sum \gamma_i t q_{i,t} - \sum_j \left[ \zeta_{j,t}(\beta, \gamma) + w_j \left( \sum_{i \in I_j} \gamma_i t \right) \right] u_{j,t} \tag{21}
$$

subject to

$$u_{j,t} - q_{i,t} \leq 0 \quad \forall i \in I_j, \forall j \tag{22}$$

$$q_{i,t} \leq 1 \quad \forall i \tag{23}$$

$$u_{j,t} \geq 0, \text{ integer} \quad \forall j. \tag{24}$$

Note that the first constraint ensures that a product is open only if all the resources it uses are open.

Now, observe that the constraint matrix of the above integer program has exactly one $+1$ and one $-1$ coefficient in each row, and hence is totally unimodular. So we can ignore the integer restriction and solve (21)–(24) exactly as an LP. In fact, problem (21)–(24) can also be solved combinatorially as a flow problem: the dual of the LP can be transformed to be a network flow problem on a bipartite graph with one set of nodes representing products and the other side resources and edges representing product-resource incidence, and flow from a source to a sink node, each connected to the product and resource nodes respectively; fast algorithms of Ahuja, Orlin, Stein, and Tarjan [1] can then be used to solve the problem in time $O(|I||J| + \min(|I| |J|^2, |J|^{3/2}\sqrt{|E|}))$ where $|I|$ is the number of resources and $|E|$ is the number of edges in this graph. Therefore, problem (21)–(24) can be solved efficiently and separating the $wAR$ constraints is tractable.

We next show that $wAR$ can be formulated as a compact LP eliminating the need for generating constraints on the fly. Since the separation problem can be solved as a LP where all the fixed values $(\beta, \gamma)$ appear in the objective function only, we can fold it into the original LP as follows: First take the dual of (21)–(24) with dual variables $\pi_{i,j,t}$ corresponding to (22), and $\psi_{i,t}$ to (23):

$$
\Pi_t^{wAR}(\beta, \gamma) = \min_{\pi, \psi} \sum_i \psi_{i,t} \tag{25}
$$

subject to

$$\sum_{i \in I_j} \pi_{i,j,t} \geq - \left[ \zeta_{j,t}(\beta, \gamma) + w_j \left( \sum_{i \in I_j} \gamma_i t \right) \right] \quad \forall j \tag{26}$$

$$- \sum_{j \in J_i} \pi_{i,j,t} + \psi_{i,t} = -\gamma_{i,t} \quad \forall i \tag{27}$$

$$\pi_{i,j,t}, \psi_{i,t} \geq 0 \quad \forall i, j \in J_i. \tag{28}$$

Note that the first constraint ensures that a product is open only if all the resources it uses are open.
Then use the second constraint in the above LP to eliminate the variable \( \psi_{i,t} \) to write the dual as

\[
\Pi_t^{wAR}(\beta, \gamma) = \min_{\pi} \sum_i \left[ \sum_{j \in J_i} \pi_{i,j,t} - \gamma_{i,t} \right]
\]

s.t. \( \sum_{i \in I_j} \pi_{i,j,t} \geq - \left[ \zeta_{j,t}(\beta, \gamma) + w_j \left( \sum_{i \in I_j} \gamma_{i,t} \right) \right] \quad \forall j \) \tag{25}

\[
\sum_{j \in J_i} \pi_{i,j,t} \geq \gamma_{i,t} \quad \forall i \tag{26}
\]

Now we fold in the above LP formulation of \( \Pi_t^{wAR}(\beta, \gamma) \) into constraints (19) and write \( wAR \) equivalently as

\[
V^{wAR} = \min_{\beta, \gamma, \pi} \sum_t \beta_t + \sum_t \sum_i \gamma_i \pi_i^1
\]

s.t. \( \beta_t \geq \sum_i \left[ \sum_{j \in J_i} \pi_{i,j,t} - \gamma_{i,t} \right] \quad \forall t \)

\tag{25}, (26) \forall t

\[
\gamma_{i,t}, \pi_{i,j,t} \geq 0 \quad \forall i, j \in J_i, t.
\]

The size of the above LP is polynomial in the number of resources and products. Hence, not only is \( wAR \) stronger than \( CDLP \), it is also tractable and has a compact formulation. Notice that this formulation would have been hard to derive and justify without the line of reasoning starting from \( AF \).

The dual of the above LP gives more insight into the weak affine relaxation. We get the dual LP as

\[
V^{wAR} = \max_{x, \rho} \sum_t \sum_j \lambda_j f_j x_{j,t}
\]

\[\text{(dwAR) s.t.} \]

\[
x_{0,t} + \sum_{s=1}^{t-1} \sum_{j \in J_i} \lambda_j x_{j,s} + \sum_{j \in J_i} x_{j,t} - \rho_{i,t} \leq r_i^1 \quad \forall i, t
\]

\[
x_{0,t} + \sum_j x_{j,t} = 1 \quad \forall t
\]

\[
\frac{x_{j,t}}{w_j} - x_{0,t} + \rho_{i,t} \leq 0 \quad \forall i, j \in J_i, t
\]

\[
x_{0,t}, x_{j,t}, \rho_{i,t} \geq 0 \quad \forall i, j, t.
\]

If we interpret \( x_{j,t} \) as the sales rate for product \( j \) at time \( t \) and \( x_{0,t} - \rho_{i,t} \) as the resource level no-purchase rate at time \( t \), then we can view \( wAR \) as a refinement of \( SBLP \) of Gallego et al. [5], where
the sales rates at each time period are modulated by the expected remaining resource capacities.

5 Tighter, tractable relaxations

The weak affine relaxation is based on isolating the difficult term in the affine relaxation and replacing it with a simpler, more tractable term. In this section, we build on this idea and propose two tractable approximation methods that further tighten the \( wAR \) bound. We again restrict attention to the single-segment MNL model to reduce notational overhead. In §6, we describe extensions to multi-segment variants of the MNL model.

5.1 Weak affine relaxation\(^+\) (\( wAR^+ \))

In this section we describe a simple way to tighten the \( wAR \) bound while retaining a compact formulation. Associating decision variables \( q_{i,t} \) and \( u_{j,t} \), respectively, to indicate if resource \( i \) and product \( j \) are open, the \( AF \) separation problem (15) can be written as

\[
\Pi_t^{AF}(\beta, \gamma) = \max_{q, u} \quad -\sum_i \gamma_{i,t} q_{i,t} \left( 1 + \sum_j w_j u_{j,t} \right) - \sum_j \zeta_{j,t}(\beta, \gamma) u_{j,t} \\
\text{s.t.} \quad (22), (23), (24).
\]

Now \( wAR \) replaces the product term \( q_{i,t} u_{j,t} \) for \( j \notin J_i \) in the first summation with 0 and since \( q_{i,t} u_{j,t} \geq 0 \), we have \( \Pi_t^{AF}(\beta, \gamma) \leq \Pi_t^{wAR}(\beta, \gamma) \). Noting that \( q_{i,t} u_{j,t} \geq q_{i,t} + u_{j,t} - 1 \), we propose replacing the right hand side of constraints (16) with

\[
\Pi_t^{wAR^+}(\beta, \gamma) = \max_{q, u} \quad -\sum_i \gamma_{i,t} q_{i,t} - \sum_{j \notin J_i} \gamma_{i,t} w_j u_{j,t} - \sum_j \left[ \zeta_{j,t}(\beta, \gamma) + w_j \left( \sum_{i \in J_j} \gamma_{i,t} \right) \right] u_{j,t} \\
\text{s.t.} \quad (22), (23) \\
\chi_{i,j,t} \geq q_{i,t} + u_{j,t} - 1 \quad \forall i, j \notin i \\
u_{j,t}, \chi_{i,j,t} \geq 0 \quad \forall j, i \notin J_j.
\]

The following lemma is immediate.

Lemma 2. \( \Pi_t^{AF}(\beta, \gamma) \leq \Pi_t^{wAR^+}(\beta, \gamma) \leq \Pi_t^{wAR}(\beta, \gamma) \).
Therefore, we replace the right hand side of constraints (16) with $\Pi^{WAR^+}(\beta, \gamma)$ and solve the LP

\[
V^{WAR^+} = \min_{\beta, \gamma} \sum_t \beta_t + \sum_t \sum_i \gamma_{it} r_{i1}^t \\
(\text{WAR}^+) \text{ s.t. } \beta_t \geq \Pi^{WAR^+}(\beta, \gamma) \forall t \\
\gamma_{it} \geq 0 \forall i, t.
\]  

We refer to this method as weak affine relaxation+ ($\text{WAR}^+$). Lemma 2 together with Lemma 1 implies that $V^{AF} \leq V^{WAR^+} \leq V^{WAR}$. Therefore, $\text{WAR}^+$ further tightens the $\text{WAR}$ bound. Note however that the $\text{WAR}^+$ separation problem can have as many as $|\mathcal{I}| |\mathcal{J}|$ additional constraints compared to $\text{WAR}$. Still, the $\text{WAR}^+$ separation problem involves solving an LP and hence is tractable. Moreover, it is possible to obtain a compact formulation of $\text{WAR}^+$ by following the steps in §4.3; we omit the details.

5.2 A hierarchical family of relaxations

In this section we show how to construct a hierarchical family of relaxations that at the highest level (level-\(n\), the number of products) gives us the affine relaxation. Naturally, because of the NP-hardness of solving the affine relaxation, we cannot expect tractability, and so we concentrate on small levels. The level-1 relaxation already turns out to be a tighter relaxation than $\text{WAR}$. While the level-1 relaxation separation problem can be solved in a tractable manner, a potential drawback is that, unlike $\text{WAR}$ and $\text{WAR}^+$, it cannot be folded into the original problem to yield a compact formulation.

For simplicity we describe the level-1 formulation and remark on how it extends to a hierarchy of relaxations. In the level-1 relaxation, which we refer to as hierarchical affine relaxation ($\text{hAR}$), we replace the $\gamma_{S,t}(1 + \sum_{j \in S} w_j)$ term in (15) with $\gamma_{S,t} + (\sum_{j \in S} w_j)(\max_{j' \in S} \sum_{i \in I_{j'}} \gamma_{i,t})$ and solve the LP

\[
V^{hAR} = \min_{\beta, \gamma} \sum_t \beta_t + \sum_t \sum_i \gamma_{it} r_{i1}^t \\
(\text{hAR}) \text{ s.t. } \beta_t \geq \Pi^{hAR}(\beta, \gamma) \forall t \\
\gamma_{it} \geq 0 \forall i, t.
\]
where
\[
\Pi_t^{hAR} = \max_S \left\{ -\gamma_{S,t} - \left( \sum_{j \in S} w_j \right) \left( \max_{j' \in S} \sum_{i \in I_{j'}} \gamma_{i,t} \right) - \sum_{j \in S} \zeta_{j,t}(\beta, \gamma) \right\}.
\]

We have the following lemma.

**Lemma 3.** $\Pi_t^{AF}(\beta, \gamma) \leq \Pi_t^{hAR}(\beta, \gamma) \leq \Pi_t^{wAR}(\beta, \gamma)$.

**Proof**

By definition, we have $I_{\hat{j}} \subseteq I_S$ for all $j \in S$. Therefore, $\gamma_{S,t} = \sum_{i \in I_S} \gamma_{i,t} \geq \sum_{i \in I_{\hat{j}}} \gamma_{i,t}$ for all $j \in S$ and so $\gamma_{S,t} \geq \max_{j \in S} \sum_{i \in I_{j'}} \gamma_{i,t}$. The proof now follows by noting that $\gamma_{S,t} \left( 1 + \sum_{j \in S} w_j \right) \geq \gamma_{S,t} + \left( \sum_{j \in S} w_j \right) \left( \max_{j' \in S} \sum_{i \in I_{j'}} \gamma_{i,t} \right) \geq \gamma_{S,t} + \sum_{j \in S} w_j \left( \sum_{i \in I_{j}} \gamma_{i,t} \right).

Lemma 3 together with Lemma 1 implies that $V^{AF} \leq V^{hAR} \leq V^{wAR}$. Therefore, $hAR$ obtains a tighter bound than $wAR$.

Next, we show that $hAR$ separation problem (29) can be solved in a tractable manner. Associating binary decision variables $q_{i,t}$ and $u_{j,t}$, respectively, to indicate if resource $i$ and product $j$ are open, problem (29) can be written as

\[
\Pi_t^{hAR}(\beta, \gamma) = \max_{q,u} \quad -\sum_{i} \gamma_{i,t} q_{i,t} - \left( \sum_{j} w_j u_{j,t} \right) \left( \max_{j'} \sum_{i \in I_{j'}} \gamma_{i,t} \right) - \sum_{j} \zeta_{j,t}(\beta, \gamma) u_{j,t}
\]

\[\text{s.t. (22) - (24).}\]

Although the above optimization problem has a nonlinear objective function, we can solve it through a sequence of LPs in the following manner. We fix a product $\hat{j}$ as the one achieving the maximum value of $\max_{j} \gamma_{i,t} u_{j,t}$. Since $\hat{j}$ achieves the maximum value, we must have $u_{\hat{j},t} = 1$ and $u_{j,t} = 0$ for $j$ with $\sum_{i \in I_{\hat{j}}} \gamma_{i,t} > \sum_{i \in I_{j}} \gamma_{i,t}$. Letting $\hat{J} = \left\{ j \mid \sum_{i \in I_{j}} \gamma_{i,t} > \sum_{i \in I_{\hat{j}}} \gamma_{i,t} \right\}$, we solve the following linear integer program for product $\hat{j}$:

\[
\Pi_t^{hAR}(\beta, \gamma) = \max_{q,u} \quad -\sum_{i} \gamma_{i,t} q_{i,t} - \sum_{j} \zeta_{j,t}(\beta, \gamma) \left( \sum_{i \in I_{j}} \gamma_{i,t} \right) u_{j,t}
\]

\[\text{s.t. (22), (23)} \quad u_{\hat{j},t} = 1 \quad u_{j,t} = 0 \quad \forall j \in \hat{J} \quad u_{j,t} \geq 0 \text{ integer} \quad \forall j \in J \setminus \hat{J}.\]
Since the constraint matrix is totally unimodular, we can solve the above linear integer program equivalently as an LP. So we solve the LP for each product $\tilde{j} \in J$ and obtain $\Pi_t^{hAR}(\beta, \gamma) = \max_{\tilde{j} \in J} \Pi_t^{hAR, \tilde{j}}(\beta, \gamma)$.

Since problem (29) can be solved in a tractable manner, separating the $hAR$ constraints is tractable, and $hAR$ can be solved in polynomial time by the ellipsoid method. However, unlike $wAR^+$, $hAR$ does not seem to have a compact LP formulation. This is because the set $\tilde{J}_j$ depends on the values of the $\gamma$'s in a nonlinear fashion and the duality argument in §4.3 that we used to fold the separation problem back into the original LP does not hold. On the other hand, an appealing feature of $hAR$ is that its separation problem has fewer number of decision variables and constraints than $wAR^+$.

Remark: One can get further relaxations by considering pairs of elements $j', j''$ for a level-2 relaxation (or triples for level-3, and so on) such that we find the offer set $S$ that maximizes

$$- \left( 1 + \sum_{j \in S} w_j \right) \left[ \max_{(j', j'' \in S)} \sum_{i \in I(j', j'')} \gamma_{i,t} \right].$$

In this way, we can control the degree of approximation to the affine relaxation. We limit our numerical results to fixing a single element $j'$.

6 MNL with multiple customer segments

In this section we describe how to extend the weak affine relaxation of §4 to the mixture-of-multinomial-logits (MMNL) model (the development for $wAR^+$ and $hAR$ is similar). The MMNL model is a rich choice model that can approximate any random utility choice model arbitrarily closely; McFadden and Train [8]. In §6.1 we consider the MMNL choice model with disjoint consideration sets. In §6.2 we consider the case where the consideration sets of the different segments overlap. It is also possible to extend the weak affine relaxation idea to the general attraction model of Gallego et al. [5] in a transparent manner.

6.1 Disjoint consideration sets

We consider the case where the total demand is comprised of demand from multiple customer segments. The consideration sets of the different segments are disjoint and so we have $C_i \cap C_m = \emptyset$
for segments \( l \) and \( m \). We note that the case of disjoint consideration sets for the segments is one of the few known cases where the CDLP formulation is tractable. We describe below how \( wAR \) can be extended to tighten the CDLP bound in a tractable manner. The key idea is to look at the \( AF \) separation problem for each customer segment, which again turns out to be intractable. We apply the ideas from the single-segment case to get a tractable relaxation.

Let \( I_l = \{ i \in I \mid \exists j \in C_l \text{ and } j \in J_i \} \) and \( L_i = \{ l \in L \mid i \in I_l \} \). We can interpret \( I_l \) as the set of resources that are used by segment \( l \) and \( L_i \) as the set of segments that use resource \( i \). Letting \( \lambda_l \) denote the arrival rate for segment \( l \), we can interpret \( \sum_{l \in L_i} \lambda_l \) as the effective arrival rate for resource \( i \).

Now consider the separation problem for \( AF \). Using \( \lambda Q_i(S) = \sum_l \lambda_l Q^l_i(S_l) \) and \( \lambda R(S) = \sum_l \lambda_l R^l(S_l) \), where \( S_l = S \cap C_l \), constraint (12) can be written as

\[
\beta_t + \sum_i \sum_{l \in L_i} \gamma_{i,t} \sum_{l \in L_i} \lambda_l \left[ \left( \sum_{k=t+1}^\tau \gamma_{i,k} \right) \sum_l \lambda_l Q^l_i(S_l) \right] \geq \sum_l \lambda_l R^l(S_l). \tag{30}
\]

We first split this constraint into \( l \) separate constraints, one for each segment, by introducing variables \( \beta_{l,t} \). The constraint for segment \( l \) at time \( t \) is that

\[
\beta_{l,t} + \sum_i \sum_{l \in L_i} \gamma_{i,t} \sum_{l \in L_i} \lambda_l \left[ \left( \sum_{k=t+1}^\tau \gamma_{i,k} \right) \lambda_l Q^l_i(S_l) \right] \geq \lambda_l R^l(S_l). \tag{31}
\]

for each \( S_l = S \cap C_l \), where \( \lambda_l / \sum_{l' \in L_i} \lambda_{l'} \) can be interpreted as the probability of a segment-\( l \) arrival given the arrival of a segment that uses resource \( i \). The proof of Proposition 4 below shows that the segment level constraints (31) imply (30) and that we obtain a looser upper bound by separating over (31) instead of (30).

We observe that the segment level constraints (31) have the same form as constraints (12) in the single-segment case, and are therefore hard to separate. So we use the same relaxation as we did for the single-segment case to obtain a tractable separation problem at the segment level:

\[
\Pi_{t,l}^{swAR}(\beta, \gamma) = \max_{q,u} - \sum_{i \in I_l} \sum_{l' \in L_i} \frac{\lambda_l \gamma_{i,t}}{\lambda_{l'}} q_{i,t} - \sum_{j \in L_i} w_j \left[ \beta_{l,t} + \lambda_l \left( \sum_{i \in I_j} \sum_{k=t+1}^\tau \gamma_{i,k} \right) - f_j + \sum_{l' \in L_i} \frac{\gamma_{i,t}}{\lambda_{l'}} \right] u_{j,t} \tag{22} - \tag{24}.
\]

We replace constraint (31) with \( \beta_{l,t} \geq \Pi_{t,l}^{swAR}(\beta, \gamma) \) to obtain a segment-based weak affine relaxation.
(swAR):

\[
V^{\text{swAR}} = \min_{\beta, \gamma} \sum_t \sum_l \beta_{l,t} + \sum_t \sum_i \gamma_{i,t} r_i^1 \\
\text{s.t.} \quad \beta_{l,t} \geq \Pi^{\text{swAR}}_{l,t}(\beta, \gamma) \quad \forall l, t \\
\gamma_{i,t} \geq 0 \quad \forall i, t.
\]

Moreover, by following the same steps as for the single-segment case, it is possible to show that swAR can be formulated as the following compact LP

\[
V^{\text{swAR}} = \min_{\beta, \gamma, \tau} \sum_t \sum_l \beta_{l,t} + \sum_t \sum_i \gamma_{i,t} r_i^1 \\
\text{s.t.} \quad \beta_{l,t} \geq \sum_{i \in I_i, j \in C_i} \pi_{i,j,t} - \frac{\lambda_i}{\sum_{l' \in L_i} \lambda_{l'}} \gamma_{i,t} \quad \forall l, t \\
\sum_{i \in I_j} \pi_{i,j,t} \geq \lambda_{\ell_j} w_{\ell_j}^j \left[ f_j - \sum_{i \in I_{j', i \in C_{j'}}} \left( \gamma_{i,t} + \frac{\gamma_{i,t}}{\sum_{l' \in L_i} \lambda_{l'}} \right) - \frac{\beta_{l', t}}{\lambda_{l'}} \right] \quad \forall j, t \\
\sum_{j \in J_i, j \in C_i} \pi_{i,j,t} \geq 0 \quad \forall i, l \in L_i, t \\
\gamma_{i,t}, \pi_{i,j,t} \geq 0 \quad \forall i, j \in J_i, t,
\]

where \(\ell_j\) denotes the segment to which product \(j\) belongs. swAR can be viewed as an extension of wAR to the MNL model with multiple segments and disjoint consideration sets. In particular, swAR coincides with wAR if there is only a single segment. Note that swAR is again tractable as it is a compact LP. Proposition 4 below shows that it also obtains an upper bound on the value function that is tighter than CDLP.

**Proposition 4.** \(V^{AF} \leq V^{\text{swAR}} \leq V^{CDLP}\).

**Proof**

In Online Appendix.

As we show in the next section, it is possible to extend the swAR formulation to the MNL model with multiple segments when the consideration sets overlap. The dual of swAR, which we give
below, turns out to be useful for this purpose.

\[
V^{dswAR} = \max_{x, \rho} \sum_{l} \sum_{t} \sum_{j \in C_l} \lambda_l f_j x_{j,t}^{l} \\
\text{(dswAR) s.t.} \sum_{l \in L} \lambda_l \left( \frac{x_{0,t}^{l} + \sum_{s=1}^{t-1} x_{j,s}^{l} + \sum_{j \in J_l \cap C_l} x_{j,t}^{l} \lambda_{l'} - \frac{\rho_{l,t}^{i}}{\sum_{l' \in L} \lambda_{l'}}}{\sum_{l' \in L} \lambda_{l'}} \right) \leq r_i^{l} \quad \forall i, t \quad (32)
\]

\[
x_{0,t}^{l} + \sum_{j \in C_l} x_{j,t}^{l} = 1 \quad \forall l, t
\]

\[
x_{j,t}^{l} \frac{w_j}{w_j} - x_{0,t}^{l} + \rho_{l,t}^{i} \leq 0 \quad \forall l, i, j \in J_l \cap C_l, t \quad (33)
\]

\[
x_{0,t}^{l}, x_{j,t}^{l}, \rho_{l,t}^{i} \geq 0 \quad \forall l, i, j \in J_l \cap C_l, t.
\]

6.2 Overlapping consideration sets

When the segment consideration sets overlap, the CDLP formulation is difficult to solve, even for MNL with just two segments. So one would imagine that it is difficult to find a tractable bound tighter than CDLP in this case. One strategy, pursued in Meissner, Strauss, and Talluri [10] is to formulate the problem by segments and then add a set of consistency conditions called product-cut equalities (PC-equalities). These equalities apply to any general discrete-choice model and appear to be quite powerful in numerical experiments, often bringing the solution close to CDLP value. Strauss and Talluri [13] subsequently show that when the consideration set structure has a certain tree structure, the cuts in fact achieve the CDLP value. Talluri [14] shows how to specialize the PC-equalities to the MNL choice model. In this section we describe how the PC-equalities, specialized for MNL, can be added to dswAR to tighten the formulation.

We begin with a brief description of the PC-equalities: Meissner et al. [10] allow different sets to be offered to different segments. However, to ensure consistency, they require that for any product \( j \in C_l \cap C_m \), the length of time it is offered to segment \( l \) must be the equal to the length of time it is offered to segment \( m \). This leads to a set of consistency constraints which they term as PC-equalities.
Talluri [14] uses choice probabilities (1) to specialize the PC-equalities to the MNL model as:

\[
x^j_l = \sum_{\{S \subseteq (C_l \cap C_m) | j \in S\}} \frac{y^l_{S,m}}{w^l_j} \quad \forall l, m, j \in C_l \cap C_m
\]

(34)

\[
y^l_{S,j} \leq y^l_{S,m} \quad \forall l, m, S \subseteq C_l \cap C_m, j \in C_l \setminus C_m
\]

(35)

\[
\sum_{\{T \subseteq (C_l \cap C_m) | T \supseteq S\}} \left\{ \sum_{j \in C_l \setminus C_m} w^l_j y^l_{T,j} + (1 + W^l_T)y^l_{T,m} \right\} = \sum_{\{T' \subseteq (C_m \cap C_l) | T' \supseteq S\}} \left\{ \sum_{j \in C_m \setminus C_l} w^m_j y^m_{T',j} + (1 + W^m_{T'})y^m_{T',m} \right\} \quad \forall l, m, S \subseteq C_l \cap C_m,
\]

(36)

where \( W^l_T = \sum_{j \in S} w^l_j \) and we have new variables of the form \( y^l_{S,m} \) defined for all pairs of segments \( l, m \) and for all \( S \subseteq C_l \cap C_m \); see Talluri [14]. If the overlap in the consideration sets of the different segments is not too large, then the number of PC-equalities is manageable.

Talluri [14] shows that adding PC-equalities (34)-(36) to the sales-based linear program (SBLP) of Gallego et al. [5] further tightens the SBLP bound. We are also able to tighten the \( dswAR \) bound by doing the same thing. Moreover, comparing \( dswAR \) with \( SBLP \), it is easy to see that a feasible solution to \( dswAR \) is also feasible to \( SBLP \). Therefore, \( dswAR \) is tighter than \( SBLP \).

It follows that \( dswAR \) augmented with the PC-equalities continues to be tighter than \( SBLP \) with the same PC-equalities. So in conclusion, when segment consideration sets overlap, we also have

**Proposition 5.** The objective function value of \( dswAR \) with (34–36) added, is less than or equal to the objective function value of \( SBLP \) with (34–36) added.

In closing, we note that \( dswAR \) augmented with the PC-equalities is not guaranteed to be tighter than \( CDLP \). We numerically compare the performance of \( dswAR \) with \( CDLP \) in our computational experiments that we present next.

### 7 Computational experiments

In this section, we compare the upper bounds and the revenues obtained by \( CDLP \), \( wAR \), \( wAR^+ \), \( hAR \) and \( AF \). We begin by describing the experimental setup.
7.1 Test network

We consider a hub-and-spoke network with a single hub that serves $N$ spokes. Half of the spokes have two flights to the hub, while the remaining half have two flights from the hub so that the total number of flights is $2N$ (an Online Appendix contains a graphic of the network).

The total number of fare-products is $2N(N+2)$. There are $4N$ fare products connecting spoke-to-hub and hub-to-spoke origin-destination pairs, of which half are high fare-products and the remaining half are low-fare products. The high fare-product is 50% more expensive than the corresponding low fare-product. The remaining $4N^2$ fare-products connect spoke-to-spoke origin-destination pairs. Half of the $4N^2$ fare-products are high fare-products and the rest are low fare-products, with the high fare-product being 50% more expensive than the corresponding low fare-product.

In all of our test problems, we have multiple customer segments and choice within each segment is governed by the MNL model. For each customer segment, we sample the preference weights of the fare-products in its consideration set from a poisson distribution with a mean of 100 and set the no-purchase preference weight to be $0.5 \sum_{j \in C_i} w_j$. So the probability that a customer does not purchase anything when all the products in the consideration set are offered is $1/3$.

We measure the tightness of the leg capacities using the nominal load factor, which is defined in the following manner. Letting $\hat{S}_t = \arg \max_S R(S)$ denote the optimal set of products offered at time period $t$ when there is ample capacity on all night legs, we define the nominal load factor

$$\alpha = \frac{\sum_i \lambda Q_i(\hat{S}_t)}{\sum_i \tau_i}.$$ 

We set the arrival rate $\lambda = 0.9$ and have $\tau = 200$ time periods in all of our test problems. We label our test problems by $(N, \alpha)$ where $N \in \{4, 6, 8\}$ and $\alpha \in \{0.8, 1.0, 1.2, 1.6\}$.

7.2 Results

We first consider test problems where the consideration sets of the different customer segments are disjoint. In these test problems, each origin-destination pair is associated with a customer segment and the segment is only interested in the fare-products connecting that particular origin-destination pair.

Table 1 gives the upper bounds obtained by the benchmark solution methods. The first column in the table gives the problem characteristics. The second to sixth columns, respectively, give the
upper bounds obtained by $CDLP$, $wAR$, $wAR^+$, $hAR$, and $AF$. The last four columns give the percentage gap between the upper bounds obtained by $CDLP$ and $wAR$, $CDLP$ and $wAR^+$, $CDLP$ and $hAR$, and $CDLP$ and $AF$, respectively. We note that $wAR$, $wAR^+$ and $hAR$ correspond to the segment-based versions of the relaxations described in §6 and by $AF$, we mean the reduced formulation $RAF$.

$AF$ generates the tightest upper bound and $CDLP$ the weakest, with the remaining upper bounds sandwiched in between. The average percentage gap between $wAR$ and $CDLP$ is 1.59%, although we observe instances where the gap is as high as 2.73%. The percentage gap between $wAR$ and $CDLP$ seems to increase with the nominal load factor and the number of spokes in the network. $wAR^+$ and $hAR$ tighten the $wAR$ bound and obtain bounds that are on average 1.81% and 1.63% tighter than $CDLP$. $AF$ obtains bounds that are on average 2.16% tighter than $CDLP$. $wAR$ closes about 70% of the gap between the $AF$ and $CDLP$ bounds; the corresponding numbers for $wAR^+$ and $hAR$ are 80% and 75%, respectively.

Table 2 gives the expected revenues obtained by the different benchmark methods. We evaluate the revenue performance by simulation and use common random numbers in our simulations. In our revenue simulations, we divide the booking period into five equal intervals. At the beginning of each interval, we re-solve the benchmark solution methods to get fresh estimates for the marginal value.
Table 2: Comparison of the expected revenues for the hub-and-spoke test problems with disjoint consideration sets.

<table>
<thead>
<tr>
<th>Problem</th>
<th>Expected Revenue</th>
<th>% Gap with CDLP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CDLP</td>
<td>wAR</td>
</tr>
<tr>
<td>(4, 0.8)</td>
<td>5,755</td>
<td>5,748</td>
</tr>
<tr>
<td>(4, 1.0)</td>
<td>5,263</td>
<td>5,242</td>
</tr>
<tr>
<td>(4, 1.2)</td>
<td>5,056</td>
<td>5,080</td>
</tr>
<tr>
<td>(4, 1.6)</td>
<td>4,413</td>
<td>4,570</td>
</tr>
<tr>
<td>(6, 0.8)</td>
<td>5,487</td>
<td>5,531</td>
</tr>
<tr>
<td>(6, 1.0)</td>
<td>5,047</td>
<td>5,127</td>
</tr>
<tr>
<td>(6, 1.2)</td>
<td>4,665</td>
<td>4,764</td>
</tr>
<tr>
<td>(6, 1.6)</td>
<td>3,824</td>
<td>4,101</td>
</tr>
<tr>
<td>(8, 0.8)</td>
<td>4,829</td>
<td>4,888</td>
</tr>
<tr>
<td>(8, 1.0)</td>
<td>4,343</td>
<td>4,434</td>
</tr>
<tr>
<td>(8, 1.2)</td>
<td>3,969</td>
<td>4,091</td>
</tr>
<tr>
<td>(8, 1.6)</td>
<td>3,384</td>
<td>3,579</td>
</tr>
</tbody>
</table>

avg. 2.28 3.02 2.94 2.34

The columns in Table 2 have a similar interpretation as in Table 1 except that they give the expected total revenues. In the last four columns, we use ✓ to indicate that the corresponding benchmark method generates higher revenues than CDLP at the 95% level, an ⊙ if the difference in the revenue performance of the benchmark method and CDLP is not significant at the 95% level and a × if the benchmark method generates lower revenues than CDLP at the 95% level. wAR on average generates revenues that are 2.28% higher than CDLP, although we observe instances where the gap is as high as 7%. wAR+, hAR and AF, respectively, generate revenues that are on average 3.02%, 2.94% and 2.34%, higher than the CDLP revenues.

Table 3 shows how the CPU seconds required by the different solution methods vary with the number of spokes in the network. All of our computational experiments are carried out on a Pentium Core 2 Duo desktop with 3-GHz CPU and 4-GB RAM. We use CPLEX 11.2 to solve all LPs. We solve wAR and wAR+ to optimality since they have compact formulations. For MMNL with disjoint
consideration sets, CDLP has a compact sales-based formulation (SBLP) as well and can be solved to optimality. On the other hand, we solve hAR and AF by generating constraints on the fly and stop when we are within 1% of optimality.

CDLP can be solved in a matter of seconds. The solution times of the other methods are generally in minutes. wAR solves around five to ten times faster than AF and the savings can be significant especially for relatively large networks. In light of the hardness result in Proposition 2, we only expect the savings in run times to increase with the problem size. wAR and hAR have additional computational overheads associated with them and thus take longer than wAR.

Among wAR, wAR+ and hAR, wAR emerges as one that achieves a good balance between the quality of the solution and the computational effort. Going forward, we focus on comparing the performance of wAR with that of CDLP and AF.

<table>
<thead>
<tr>
<th>No. of spokes</th>
<th>CPU secs.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CDLP</td>
</tr>
<tr>
<td>6</td>
<td>0.4</td>
</tr>
<tr>
<td>8</td>
<td>0.8</td>
</tr>
<tr>
<td>10</td>
<td>1.2</td>
</tr>
<tr>
<td>12</td>
<td>1.9</td>
</tr>
</tbody>
</table>

Table 3: CPU seconds for the benchmark solution methods as a function of the number of spokes in the airline network for the test problems with disjoint consideration sets. Both hAR and AF (solved via the RAF version) are stopped when they reach 1% of optimality.

One interesting observation is that the improvements in the revenue performance of wAR (over CDLP) tend to be larger than the corresponding improvements in the upper bound. It turns out that wAR obtains sharper value function approximations closer to the end of the sales horizon when resource capacities are relatively scarce. Consequently it is able to make better decisions on the set of products to offer in capacity constrained scenarios and this yields significant benefits. We describe an additional set of computational experiments in the Online Appendix to support the above observations.

Next we consider test problems where the consideration sets of the segments overlap. We continue to work with the hub-and-spoke network test problems, except that now each origin-destination pair is associated with two customer segments. The first segment considers only the low fare products connecting the origin-destination pair, while the second segment considers the high fare products as
Table 4 gives the upper bounds obtained by CDLP, wAR and AF. Note that by wAR, we mean the segment-based weak affine relaxation augmented with product-cut equalities described in §6.2. As mentioned, when the consideration sets overlap, wAR is not provably tighter than CDLP. However, we observe that overall wAR tends to obtain tighter bounds than CDLP and the gaps can be as high as 3%.

Table 5 gives the expected revenues obtained by the different solution methods. We find that wAR continues to provide noticeable revenue boosts over CDLP and its revenue performance is competitive with AF. As in the case with disjoint consideration sets, the revenue boosts are more significant at the higher load factors.

Table 6 reports the CPU seconds required by CDLP, wAR and AF for the hub-and-spoke network test problems with overlapping consideration sets. We note that CDLP does not have a compact formulation anymore and has to be solved using column generation. We report the times required to solve CDLP to within 1% of optimality. When the consideration sets overlap, the CDLP column generation problem becomes intractable. Hence its solution time increases considerably compared to the case with disjoint consideration sets. The solution times for wAR (with product-cut equalities) are now comparable to that of CDLP. Both wAR and CDLP can be solved in a matter of minutes while AF can take hours.

8 Conclusions

CDLP and the affine relaxation are two methods in the literature that give upper bounds on the value function for choice network revenue management. While CDLP is known to be tractable for the MNL model with disjoint consideration sets, we show that the affine relaxation is NP-hard even for the single-segment MNL model. Nevertheless, by analyzing the affine relaxation we obtain tractable, weaker, approximations. We show that our approximations yield upper bounds that are in between the CDLP and the affine bounds. Our relaxations retain the appeal of the formulation discovered in Gallego et al. [5] in that they involve solving compact LPs, eliminating the need for constraint or column generation. We extend our approximations to the mixture-of-multinomial-logits model with disjoint as well as with overlapping consideration sets. Our computational study indicates that our approximations typically produce upper bounds that are close to the affine bound
<table>
<thead>
<tr>
<th>Problem (N, α)</th>
<th>Upper Bound CDLP</th>
<th>wAR AF/RAF</th>
<th>% Gap with CDLP wAR AF/RAF</th>
</tr>
</thead>
<tbody>
<tr>
<td>(4, 0.8)</td>
<td>7,069</td>
<td>7,094</td>
<td>7,060</td>
</tr>
<tr>
<td>(4, 1.0)</td>
<td>6,309</td>
<td>6,266</td>
<td>6,241</td>
</tr>
<tr>
<td>(4, 1.2)</td>
<td>5,975</td>
<td>5,907</td>
<td>5,879</td>
</tr>
<tr>
<td>(4, 1.6)</td>
<td>5,207</td>
<td>5,140</td>
<td>5,098</td>
</tr>
<tr>
<td>(6, 0.8)</td>
<td>6,783</td>
<td>6,807</td>
<td>6,773</td>
</tr>
<tr>
<td>(6, 1.0)</td>
<td>6,240</td>
<td>6,149</td>
<td>6,109</td>
</tr>
<tr>
<td>(6, 1.2)</td>
<td>5,789</td>
<td>5,683</td>
<td>5,645</td>
</tr>
<tr>
<td>(6, 1.6)</td>
<td>4,770</td>
<td>4,704</td>
<td>4,675</td>
</tr>
<tr>
<td>(8, 0.8)</td>
<td>6,783</td>
<td>6,439</td>
<td>6,291</td>
</tr>
<tr>
<td>(8, 1.0)</td>
<td>5,799</td>
<td>5,367</td>
<td>5,236</td>
</tr>
<tr>
<td>(8, 1.2)</td>
<td>4,109</td>
<td>4,384</td>
<td>4,370</td>
</tr>
<tr>
<td>(8, 1.6)</td>
<td>3,528</td>
<td>3,674</td>
<td>3,641</td>
</tr>
<tr>
<td>avg.</td>
<td></td>
<td></td>
<td>1.24</td>
</tr>
</tbody>
</table>

Table 4: Comparison of the upper bounds for the hub-and-spoke test problems with overlapping consideration sets.

<table>
<thead>
<tr>
<th>Problem (N, α)</th>
<th>Expected Revenue CDLP</th>
<th>wAR AF/RAF</th>
<th>% Gap with CDLP wAR AF/RAF</th>
</tr>
</thead>
<tbody>
<tr>
<td>(4, 0.8)</td>
<td>6,862</td>
<td>6,828</td>
<td>6,835</td>
</tr>
<tr>
<td>(4, 1.0)</td>
<td>5,827</td>
<td>5,887</td>
<td>5,913</td>
</tr>
<tr>
<td>(4, 1.2)</td>
<td>5,515</td>
<td>5,584</td>
<td>5,650</td>
</tr>
<tr>
<td>(4, 1.6)</td>
<td>4,592</td>
<td>4,774</td>
<td>4,750</td>
</tr>
<tr>
<td>(6, 0.8)</td>
<td>6,337</td>
<td>6,439</td>
<td>6,291</td>
</tr>
<tr>
<td>(6, 1.0)</td>
<td>5,799</td>
<td>5,738</td>
<td>5,730</td>
</tr>
<tr>
<td>(6, 1.2)</td>
<td>5,147</td>
<td>5,367</td>
<td>5,236</td>
</tr>
<tr>
<td>(6, 1.6)</td>
<td>4,109</td>
<td>4,357</td>
<td>4,390</td>
</tr>
<tr>
<td>(8, 0.8)</td>
<td>5,554</td>
<td>5,591</td>
<td>5,557</td>
</tr>
<tr>
<td>(8, 1.0)</td>
<td>4,803</td>
<td>4,894</td>
<td>4,887</td>
</tr>
<tr>
<td>(8, 1.2)</td>
<td>4,267</td>
<td>4,384</td>
<td>4,370</td>
</tr>
<tr>
<td>(8, 1.6)</td>
<td>3,528</td>
<td>3,674</td>
<td>3,641</td>
</tr>
<tr>
<td>avg.</td>
<td></td>
<td></td>
<td>2.17</td>
</tr>
</tbody>
</table>

Table 5: Comparison of the expected revenues for the hub-and-spoke test problems with overlapping consideration sets.
Table 6: CPU seconds for the benchmark solution methods as a function of the number of spokes in the airline network for the test problems with overlapping consideration sets. Both $CDLP$ and $AF$ (solved via the $RAF$ version) are stopped when they reach 1% of optimality. (reaching nearly 75% of the affine bound on average), have good revenue performance (obtaining on average above 95% of the revenue possible from policies based on the affine relaxation) and are tractable alternatives to solving the affine relaxation.

References


Online Appendix

Proof of Proposition 2

Proof
Our reduction is from the NP-complete maximum edge biclique problem (Peeters [11]). We state first the definitions and notation in the problem.

The problem is defined on an undirected, bipartite graph $G = (V_1 \cup V_2, E)$, with $|V_2| = m_2$. A $(k_1, k_2)$-biclique is a complete bipartite subgraph of $G$, i.e., a subgraph consisting of a pair $(X, Y)$ of vertex subsets $X \subseteq V_1$ and $Y \subseteq V_2$, $|X| = k_1 > 1, |Y| = k_2 > 1$, such that there exists an edge $(x, y) \in E, \forall x \in X, y \in Y$. Note that the number of edges in the biclique is $k_1k_2$.

Maximum edge biclique problem (MBP)

*Input:* A bipartite graph $G = (V_1 \cup V_2, E)$ and a positive integer $p$.

*Question:* Does $G$ contain a biclique with at least $p$ edges?

Consider the complement bipartite graph $\tilde{G}$ of $G$ defined on the same vertex set as $G$, where there is an edge $e = (u, v)$ in graph $\tilde{G}$ if and only if there is no edge between $u$ and $v$ in $G$.

Define a cover $C_S \subseteq V_2$ of a subset $S \subseteq V_1$ in the complement graph $\tilde{G}$, as $C_S = \{ v \in V_2 \mid \exists e = (u, v) \in \tilde{G}, u \in S \}$. By definition if $C_S$ is a cover of some subset $S$, it means there is no edge from any $u \in S$ to any $v \in V_2 \setminus C_S$ in the graph $\tilde{G}$. Hence, as $G$ is a complement of $\tilde{G}$, there is an edge from every $u \in S$ to every $v \in V \setminus C(S)$ in $G$, thus representing a biclique between $S$ and $V \setminus C(S)$ in the graph $G$.

Now we set up the reduction for the separation for (14). In equation (14), for each $u \in V_1$, we associate a product $j$ with $f_j = m_2 \frac{(p+1)}{p}$ and $w_j = m_2$. For each $v \in V_2$, we associate a resource $i$ with weights $\gamma_{i,t} = \frac{1}{p}$ and $\gamma_{i,k} = 0, k > t$. The resource consumptions of the products $j$ are defined from the graph $\tilde{G}$: $j$ contains all the $i$ such that there is an edge between the associated nodes in $\tilde{G}$. We let $\lambda = 1, \beta_t = m_2$.

We now claim that $G$ has a $(k_1, k_2)$-biclique with $k_1k_2 > p$ if and only if there is a set $S$ that violates the inequality (14) for this instance.
With the above values, $S \subseteq V_1$, with $|S| = k_1$, $|C(S)| = m_2 - k_2$ violates (14) if and only if

$$m_2 - \sum_{j \in S} \frac{(p+1)(m_2)^2}{p(1+\sum_{j \in S} m_2)} < -\sum_{i \in C(S)} \frac{1}{p}$$

or,

$$m_2 - \frac{(p+1)m_2k_1}{p}\left(\frac{1}{m_2} + k_1\right) < -\frac{(m_2 - k_2)}{p}$$

or multiplying both sides by the positive number $p(\frac{1}{m_2} + k_1)$,

$$m_2p\left(\frac{1}{m_2} + k_1\right) - (p+1)m_2k_1 < -(m_2 - k_2)\left(\frac{1}{m_2} + k_1\right)$$

or,

$$p < \frac{(m_2 - k_2)}{m_2} + k_2k_1.$$  

The term $0 < \frac{(m_2 - k_2)}{m_2} < 1$ implies, if and only if

$$p < k_2k_1.$$  

\[\square\]

\textbf{Proof of Proposition 4}

Proof

Using the MNL choice probability (1) and rearranging terms, the $swAR$ constraint $\beta_{l,t} \geq \Pi_{l,t}^{swAR}(\beta, \gamma)$ can be equivalently written as

$$\beta_{l,t} \geq \lambda_l \left[R^l(S_l) - \sum_{i \in I_l} \sum_{k=t+1}^r Q^l_i(S_l) \gamma_{i,k} - \sum_{i \in I_l} \Pi_{[i \in I_l]} \gamma_{i,t} \sum_{l' \in L_i} \lambda_{l'} \left(\sum_{j \in \mathcal{C}_l} P^l_j(S_l) + P^t_0(S_l)\right)\right] (37)$$

for all $S_l \subseteq \mathcal{C}_l$.

Consider now two intermediate problems:

$$\mathcal{V} = \min_{\beta, \gamma} \sum_t \sum_l \beta_{l,t} + \sum_t \sum_i \gamma_{i,t} r_{i,t}^l$$

s.t. (31) $\forall l, S_l \subseteq \mathcal{C}_l, t$

$$\gamma_{i,t} \geq 0 \quad \forall i, t.$$
and

\[
\bar{V} = \min_{\beta, \gamma} \sum_i \sum_t \beta_i,t + \sum_i \sum_{t}^{T} \gamma_i,t r_i^t \\
\text{s.t. } \beta_{i,t} \geq \lambda_i \left[ R^i(S_i) - \sum_{i \in I_t} \sum_{k=t}^{T} Q^i_t(S_i) \gamma_{i,k} \right] \quad \forall i, S_i \subseteq C_i, t
\]

(38)

We can interpret the first problem as a segment based relaxation of AF, while the second problem can be viewed as a segment based relaxation of CDLP.

We next show that \( V^{AF} \leq V \leq V^{swAR} \leq \bar{V} = V^{CDLP} \), which completes the proof of the proposition.

(i) \( V^{AF} \leq V \leq V^{swAR} \). Since the objective functions of all the problems are the same, we only need to compare the corresponding constraints. Since \( \sum_{j \in J_i} P^j_i(S_i) + P^0_i(S_i) \leq 1 \), it follows that constraint (37) implies constraint (31) and we have \( V \leq V^{swAR} \).

On the other hand, the right hand side of constraint (38) can be written as

\[
\lambda_i \left[ R^i(S_i) - \sum_{i \in I_t} \sum_{k=t}^{T} Q^i_t(S_i) \gamma_{i,k} \right] - \sum_{i \in I_t} \lambda_i Q^i_t(S_i) \gamma_{i,t}.
\]

Now note that

\[
\lambda_i Q^i_t(S_i) \gamma_{i,t} = \lambda_i \mathbb{I}_{[i \in I_t]} Q^i_t(S_i) \gamma_{i,t} = \lambda_i \mathbb{I}_{[i \in I_t]} \left[ \sum_{j \in J_i} P^j_i(S_i) \right] \gamma_{i,t}
\]

\[
\leq \frac{\lambda_i}{\sum_{j' \in J_i} \lambda_{i'}} \mathbb{I}_{[i \in I_t]} \left[ \sum_{j \in J_i} P^j_i(S_i) \right] \gamma_{i,t} \leq \frac{\lambda_i}{\sum_{j' \in J_i} \lambda_{i'}} \mathbb{I}_{[i \in I_t]} \left[ \sum_{j \in J_i} P^j_i(S_i) + P^0_i(S_i) \right] \gamma_{i,t}
\]

where the first equality holds since if \( \mathbb{I}_{[i \in I_t]} = 0 \), then \( Q^i_t(S_i) = 0 \) and the first inequality holds since \( \sum_{j' \in J_i} \lambda_{i'} \leq 1 \). Therefore constraint (38) implies constraint (37) and we have \( V^{swAR} \leq \bar{V} \).

(ii) \( V^{AF} \leq V \). Suppose that \((\hat{\beta}, \hat{\gamma})\) satisfies constraints (31). We show that it satisfies constraints
(30) as well. Fix a set \(S\) and let \(S_t = S \cap C_t\). Adding up constraints (31) for all the segments

\[
\sum_t \hat{\beta}_{t,t} \geq \sum_t \left\{ \lambda_t \left[ R^t(S_t) - \sum_{i \in \mathcal{I}_t} \sum_{k=t+1}^\tau Q^t_i(S_t) \hat{\gamma}_{i,k} \right] - \sum_{i \in \mathcal{I}_t} \mathbb{I}_{\{i \in \mathcal{I}_t\}} \frac{\lambda_t}{\sum_{i' \in \mathcal{I}_t} \lambda_{i'}} R^{i'}(S_t) \right\}
\]

\[
= \lambda \left[ R(S) - \sum_{i} \sum_{k=t+1}^\tau Q^t_i(S) \hat{\gamma}_{i,k} \right] - \sum_t \hat{\gamma}_{i,t} \sum_{i \in \mathcal{I}_t} \mathbb{I}_{\{i \in \mathcal{I}_t\}} \frac{\lambda_t}{\sum_{i' \in \mathcal{I}_t} \lambda_{i'}}
\]

\[
\geq \lambda \left[ R(S) - \sum_{i} \sum_{k=t+1}^\tau Q^t_i(S) \hat{\gamma}_{i,k} \right] - \sum_t \hat{\gamma}_{i,t} \sum_{i \in \mathcal{I}_t} \mathbb{I}_{\{i \in \mathcal{I}_t\}} \frac{\lambda_t}{\sum_{i' \in \mathcal{I}_t} \lambda_{i'}}
\]

\[
= \lambda \left[ R(S) - \sum_{i} \sum_{k=t+1}^\tau Q^t_i(S) \hat{\gamma}_{i,k} \right] - \sum_t \hat{\gamma}_{i,t} \mathbb{I}_{\{i \in \mathcal{I}_t\}},
\]

where the first equality uses the fact that \(Q^t_i(S_t) = 0\) for \(l \notin \mathcal{L}_t\) and hence \(\lambda_t Q^t_i(S_t) = \sum_{i \in \mathcal{I}_t} \lambda_t Q^t_i(S_t) = \sum_{i \in \mathcal{I}_t} \lambda_t Q^t_i(S_t)\). The second inequality holds since \(\mathbb{I}_{\{i \in \mathcal{I}_t\}} \leq \mathbb{I}_{\{i \in \mathcal{I}_t\}}\). Letting \(\tilde{\beta} = \{\tilde{\beta}_t = \sum_t \hat{\beta}_{t,t} | \forall t\}\), it follows that \((\tilde{\beta}, \hat{\gamma})\) satisfies constraints (30). Therefore \(V^\mathcal{AF} \leq \sum_t \tilde{\beta}_t + \sum_t \sum_i \hat{\gamma}_{i,t} = V\). Meissner et al. [10] prove the following that we include for completeness.

(iii) \(V = V^\mathcal{CDLP}\). (Meissner et al. [10])

Constraints (6) in \(d\mathcal{CDLP}\) are equivalent to

\[
\beta_t = \max_S \left\{ \lambda \left[ R(S) - \sum_{i} \sum_{k=t+1}^\tau Q^t_i(S) \gamma_{i,k} \right] \right\}
\]

\[
= \max_{S_t} \left\{ \sum_{i} \lambda_t \left[ R^t(S_t \cap C_t) - \sum_{i} \sum_{k=t+1}^\tau Q^t_i(S_t \cap C_t) \gamma_{i,k} \right] \right\}
\]

\[
= \sum_t \max_{S_t} \left\{ \lambda_t \left[ R^t(S_t) - \sum_{i} \sum_{k=t+1}^\tau Q^t_i(S_t) \gamma_{i,k} \right] \right\}
\]

where the last inequality uses the fact that the consideration sets are disjoint. Therefore, the \(d\mathcal{CDLP}\) constraint is equivalent to the constraints \(\beta_t = \sum_{t} \beta_{t,t}\) and

\[
\beta_{t,t} = \max_{S_t} \left\{ \lambda_t \left[ R^t(S_t) - \sum_{i} \sum_{k=t+1}^\tau Q^t_i(S_t) \gamma_{i,k} \right] \right\},
\]

which is exactly constraint (38).
Further details on computational experiments

In this section we give further details of the computational experiments. Figure 1 shows a graphic of the hub-and-spoke network we consider in our experiments.

![Figure 1: Structure of the airline network with a single hub and eight spokes.](image)

As mentioned in §7 we find that the revenue improvements of $wAR$ over $CDLP$ tend to be larger than the corresponding improvements in the upper bounds. A possible explanation is the following: $wAR$, $CDLP$ and $AF$ all yield a solution of the form $(\hat{\beta}, \hat{\gamma})$ with $\sum_{s=t}^{T} \hat{\gamma}_{i,s}$ being an estimate of the marginal value of capacity of resource $i$ at time $t$. Figure 2 shows a representative plot of how the marginal values of capacity obtained by $wAR$, $CDLP$ and $AF$ vary with time. $wAR$ and $AF$ marginal values change with time, while $CDLP$ yields static marginal values. Hence, $wAR$ does a much better job of tracking the $AF$ marginal values compared to $CDLP$, especially close to the end of the sales horizon. The effect gets magnified with the level of nonstationarity. Figure 3 shows how the marginal values of capacity obtained by $CDLP$, $wAR$ and $AF$ vary with time when we have nonstationary arrivals. Compared to stationary arrivals, $CDLP$ is a poorer approximation to $AF$. On the other hand, $wAR$ continues to do a good job of tracking the $AF$ marginal values.

Figures 2 and 3 also reveal that the marginal values of capacity obtained by $AF$ are constant for the most part and start varying only towards the end of the sales horizon. This suggests that much of the benefits of $wAR$ better tracking the $AF$ marginal values are likely to be accrued in a short time window before the end of the sales horizon when the resource capacities are relatively scarce. To verify this, we consider a fluid scaling of the hub-and-spoke test problems where we scale the flight leg capacities and the length of the sales horizon by a factor $0 < \theta \leq 1$. That is, in the $\theta$-scaled problem, the initial capacities are given by $\theta r_1$ and the length of the selling season is $\theta \tau$. 

37
Note that if we set $\theta = 1$, then we get back the original set of test problems, and smaller values of $\theta$ correspond to test problems with lower resource capacities and shorter sales horizons. Table 7 compares the upper bounds obtained by CDLP, wAR and AF as we vary the scaling parameter $\theta$. We consider the hub-and-spoke network with $N = 6$ spokes and vary $\alpha \in \{0.8, 1.0, 1.2, 1.6\}$ and $\theta \in \{1.0, 0.8, 0.6, 0.4, 0.2\}$. The first column in Table 7 gives the problem characteristics $(N, \alpha, \theta)$. The remaining columns have a similar interpretation as in Table 1. The gaps between CDLP and wAR get noticeably larger as $\theta$ becomes smaller and we observe gaps as large as 10%. Furthermore, wAR is able to close roughly 75% of the gap between CDLP and AF in such cases.

Figure 2: Marginal values of capacity obtained by CDLP, wAR and AF as a function of time for stationary arrivals. The plots are for the hub-and-spoke test problem with parameters $(6, 1.6)$. 

![Figure 2](image-url)
<table>
<thead>
<tr>
<th>Problem</th>
<th>Upper Bound</th>
<th>% Gap with CDLP</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(N, \alpha, \theta)$</td>
<td>CDLP</td>
<td>wAR</td>
</tr>
<tr>
<td>(6, 0.8, 1.0)</td>
<td>6,918</td>
<td>6,891</td>
</tr>
<tr>
<td>(6, 1.0, 1.0)</td>
<td>6,357</td>
<td>6,241</td>
</tr>
<tr>
<td>(6, 1.2, 1.0)</td>
<td>5,799</td>
<td>5,683</td>
</tr>
<tr>
<td>(6, 1.6, 1.0)</td>
<td>4,796</td>
<td>4,704</td>
</tr>
<tr>
<td>(6, 0.8, 0.8)</td>
<td>5,586</td>
<td>5,558</td>
</tr>
<tr>
<td>(6, 1.0, 0.8)</td>
<td>5,142</td>
<td>5,026</td>
</tr>
<tr>
<td>(6, 1.2, 0.8)</td>
<td>4,750</td>
<td>4,634</td>
</tr>
<tr>
<td>(6, 1.6, 0.8)</td>
<td>4,030</td>
<td>3,896</td>
</tr>
<tr>
<td>(6, 0.8, 0.6)</td>
<td>4,102</td>
<td>4,038</td>
</tr>
<tr>
<td>(6, 1.0, 0.6)</td>
<td>3,918</td>
<td>3,804</td>
</tr>
<tr>
<td>(6, 1.2, 0.6)</td>
<td>3,605</td>
<td>3,488</td>
</tr>
<tr>
<td>(6, 1.6, 0.6)</td>
<td>2,934</td>
<td>2,823</td>
</tr>
<tr>
<td>(6, 0.8, 0.4)</td>
<td>2,793</td>
<td>2,741</td>
</tr>
<tr>
<td>(6, 1.0, 0.4)</td>
<td>2,644</td>
<td>2,531</td>
</tr>
<tr>
<td>(6, 1.2, 0.4)</td>
<td>2,401</td>
<td>2,281</td>
</tr>
<tr>
<td>(6, 1.6, 0.4)</td>
<td>2,086</td>
<td>1,929</td>
</tr>
<tr>
<td>(6, 0.8, 0.2)</td>
<td>1,375</td>
<td>1,316</td>
</tr>
<tr>
<td>(6, 1.0, 0.2)</td>
<td>1,280</td>
<td>1,162</td>
</tr>
<tr>
<td>(6, 1.2, 0.2)</td>
<td>1,197</td>
<td>1,076</td>
</tr>
<tr>
<td>(6, 1.6, 0.2)</td>
<td>1,076</td>
<td>959</td>
</tr>
</tbody>
</table>

Table 7: Comparison of the upper bounds for the hub-and-spoke test problems with $N = 6$ under a fluid scaling.
Figure 3: Marginal values of capacity obtained by CDLP, wAR and AF as a function of time for nonstationary arrivals. The plots are for the hub-and-spoke test problem with parameters (6, 1.6).