Electronic Companion for
Planning Online Advertising Using Gini Indices

July 28, 2019

Miguel A. Lejeune
Department of Decision Sciences, George Washington University, Washington, DC 20052, mlejeune@gwu.edu

John Turner
The Paul Merage School of Business, University of California at Irvine, Irvine, CA 92697, john.turner@uci.edu

This e-companion accompanies the full paper of the same name, published in *Operations Research*. References for citations in this e-companion are at the end of the section where the citation is made (Appendices B, D, and E contain references).

Contents

Appendices

Appendix A: Notation, Derivations, and Proofs

A.1. Notation for Audience-Level Models & Expressions

Table EC.1 is a reference for the main notation used in the paper, which concerns models that plan blocks of impressions known as 'audience segments' or 'viewer types.'

Index	**Description**
h, i	Audience segment
j	Ad campaign
t, τ	Time period
Parameter (Single-Period Models)	**Description**
I	Set of audience segments
J	Set of ad campaigns
$\Gamma(i)$	Set of campaigns that target audience segment i
$\Gamma(j)$	Set of audience segments targeted by campaign j
Γ	Set of (audience segment, campaign) (i, j) pairs that define all campaigns' targeting requirements
d_j	Demand (in impressions) of ad campaign j
p_j	Penalty assessed for each impression of demand shortfall for campaign j
s_i	Supply (in impressions) of audience segment i
$\hat{s}_j$	Supply (in impressions) eligible for campaign j
θ_j	Demand intensity of campaign j
Decision Variable (Single-Period Models)	**Description**
x_{ij}	Proportion of the impressions of audience segment i to assign to campaign j
y_j	Demand shortfall (in impressions) for campaign j

Table EC.1 Notation

A.2. Notation for Impression-Level Models & Expressions

Table EC.2 is a reference to additional notation used in parts of this Appendix, where audience segments are treated as collections of individual impressions to express impression-level quantities for the proofs.

A.3. Full Derivation of Gini-Based Metric

We now derive our Gini-based metric (5), which measures the degree to which impressions are spread across audience segments, from first principles. We begin in the disaggregated impression-level space, where x_{rj} is the proportion of impression (arrival) r to assign to campaign j, and can be alternatively interpreted as the probability of assigning impression r to campaign j. With $\Lambda(j)$ defined as the set of impressions that match the

Index	**Description**
q, r	Impression (arrival)
Parameter (Single-Period Models)	**Description**
R	Set of impressions (arrivals)
R_i	Set of impressions that comprise audience segment i
$\Lambda(r)$	Set of campaigns that target impression r
$\Lambda(j)$	Set of impressions targeted by campaign j
Λ	Set of (impression, campaign) (r, j) pairs that define all campaigns' targeting requirements
Decision Variable (Single-Period Models)	**Description**
x_{rj}	Proportion of impression r to assign to campaign j (can be interpreted as the probability of assigning impression r to campaign j)
y_j	Demand shortfall (in impressions) for campaign j

Table EC.2 Notation for Impression-Level Expressions

targeting of campaign j, the following GMD metric, which measures how evenly-spread a campaign's ads are across impressions in a campaign's target audience, follows directly from the definition of GMD (2):

$$GMD_j = \frac{1}{|\Lambda(j)|^2} \sum_{q \in \Lambda(j)} \sum_{r \in \Lambda(j)} |x_{qj} - x_{rj}|. \tag{EC.1}$$

By aggregating impressions into mutually exclusive audience segments, and using x_{ij} as the proportion of audience segment i to assign to campaign j, we simplify equation (EC.1) as follows. For notational convenience, we order the audience segments and define $\Gamma_0(j) = \{(h, i) \in \Gamma(j)^2 : h < i\}$, which indexes all distinct audience segment pairs.

PROPOSITION EC.1. *In the aggregated audience-level space, the GMD metric corresponding to* (EC.1) *is:*

$$GMD_j = \frac{2}{\hat{s}_j^2} \sum_{(h,i) \in \Gamma_0(j)} s_h s_i |x_{hj} - x_{ij}|. \tag{EC.2}$$

Proof. Let R_i be the set of impressions represented by the audience segment i. Then we have (i) $s_i = |R_i|$, (ii) $\Lambda(j) = \cup_{i \in \Gamma(j)} R_i$, (iii) $\hat{s}_j = |\Lambda(j)| = \sum_{i \in \Gamma(j)} |R_i| = \sum_{i \in \Gamma(j)} s_i$, and (iv) $x_{rj} = x_{ij}\ \forall r \in R_i$, allowing us to simplify equation (EC.1) as follows:

$$\begin{aligned}
GMD_j &= \frac{1}{|\Lambda(j)|^2} \sum_{q \in \Lambda(j)} \sum_{r \in \Lambda(j)} |x_{qj} - x_{rj}| = \frac{1}{|\Lambda(j)|^2} \sum_{q \in \{\cup_{h \in \Gamma(j)} R_h\}} \sum_{r \in \{\cup_{i \in \Gamma(j)} R_i\}} |x_{qj} - x_{rj}| \\
&= \frac{1}{|\Lambda(j)|^2} \sum_{h \in \Gamma(j)} \sum_{i \in \Gamma(j)} \left(\sum_{q \in R_h} \sum_{r \in R_i} |x_{qj} - x_{rj}| \right) = \frac{1}{|\Lambda(j)|^2} \sum_{h \in \Gamma(j)} \sum_{i \in \Gamma(j)} \left(\sum_{q \in R_h} \sum_{r \in R_i} |x_{hj} - x_{ij}| \right) \\
&= \frac{1}{\hat{s}_j^2} \sum_{h \in \Gamma(j)} \sum_{i \in \Gamma(j)} s_h s_i |x_{hj} - x_{ij}| = \frac{2}{\hat{s}_j^2} \sum_{(h,i) \in \Gamma_0(j)} s_h s_i |x_{hj} - x_{ij}|. \qquad \square
\end{aligned}$$

The $s_h s_i$ factor enters into this expression since the number of ways we can pick a pair of impressions such that one is from audience segment h and the other is from audience segment i is precisely s_h times s_i.

Moreover, in a similar manner we can also derive the average proportion μ_j of an impression assigned to campaign j from impression-level quantities as follows:

$$\mu_j = \frac{1}{|\Lambda(j)|}\sum_{r\in\Lambda(j)} x_{rj} = \frac{1}{|\Lambda(j)|}\sum_{i\in\Gamma(j)}\sum_{r\in R_i} x_{rj} = \frac{1}{|\Lambda(j)|}\sum_{i\in\Gamma(j)}\sum_{r\in R_i} x_{ij} = \frac{1}{\hat{s}_j}\sum_{i\in\Gamma(j)} s_i x_{ij}. \tag{EC.3}$$

Finally, from (2), our Gini-based metric is $G_j = GMD_j/(2\mu_j)$.

A.4. Proof of Proposition 2: Existence of Ideal Impression Allocations

Proposition 2: Let $p_j = p, j \in J$. If the system (11a)-(11d) is feasible,

1. The optimal solution of the Gini model (SG) is always an ideal allocation. This statement is valid regardless of the emphasis placed on spreading determined by the parameter $\alpha > 0$.
2. There is no guarantee that the optimal solution of the baseline model (SB) is an ideal allocation.

Proof. Part 1) Regardless of whether perfect spreading is possible or not, the quantity pY is a valid lower bound on the optimal value of the objective function (9) of the Gini model. Any solution $\mathbf{x}$ feasible for (11a)-(11d) gives a value of (9) equal to pY and is therefore optimal. Additionally, any solution $\mathbf{x}'$ not feasible for (11a)-(11d) will either: 1) violate the minimum shortfall constraint (11b), thereby increasing the shortfall cost component in (9) without reducing the spreading cost component (i.e., since any feasible $\mathbf{x}$ has perfect spreading and minimal spreading cost equal to 0); or 2) violate the perfect spreading condition, thereby increasing the spreading cost component in (9) without reducing the shortfall cost component (i.e., since any feasible $\mathbf{x}$ has minimal spreading cost).

Part 2) Any solution $\mathbf{x}$ feasible for (11a)-(11d) minimizes the shortfall cost component in the objective function (3) of the baseline model. However, there is no guarantee that the spreading term in (3) is minimized by $\mathbf{x}$. A numerical counterexample is provided after Proposition 2. □

A.5. Proof of Proposition 3: Existence of Sufficient Orthogonality

Proposition 3:

1. The Gini objective (6) has sufficiently orthogonal spread and shortfall measures.
2. The baseline objective (3a) doesn't have sufficiently orthogonal spread and shortfall measures.

Proof. Part 1) To solve for a minimizer $\mathbf{x_j^*}$ of $f_j^{SPREAD*}(c)$, we begin by noting that the feasibility condition $f_j^{SHORTFALL}(\mathbf{x_j}) = c$ can be phrased as a restriction on the mean allocation $\mu_j = \sum_{i\in\Gamma(j)} s_i x_{ij}/\hat{s}_j$, since:

$$\begin{aligned}
&f_j^{SHORTFALL}(\mathbf{x_j}) = c \implies p_j y_j = c \implies p_j(d_j - w_j) = c\\
&\implies (d_j - w_j)/\hat{s}_j = c/(p_j\hat{s}_j) \implies (d_j/\hat{s}_j) - (w_j/\hat{s}_j) = c/(p_j\hat{s}_j)\\
&\implies \theta_j - \mu_j = c/(p_j\hat{s}_j) \implies \mu_j = \theta_j - c/(p_j\hat{s}_j).
\end{aligned}$$

Next, we claim that the solution $x_{ij}^* = \theta_j - c/(p_j\hat{s}_j)$ for all $i \in \Gamma(j)$ is a minimizer of $f_j^{SPREAD*}(c)$. By definition, the mean allocation μ_j^* under the solution x_{ij}^* is:

$$\mu_j^* = \sum_{i\in\Gamma(j)} s_i x_{ij}^*/\hat{s}_j = \sum_{i\in\Gamma(j)} s_i\left(\theta_j - c/(p_j\hat{s}_j)\right)/\hat{s}_j = \theta_j - c/(p_j\hat{s}_j),$$

which clearly satisfies $f_j^{SHORTFALL}(\mathbf{x_j}) = c$. As well, $f_j^{SPREAD*}(c) = \alpha w_j^* G_j^* = \alpha \hat{s}_j GMD_j^*/2 = 0$, where the last equality follows because $GMD_j^* = 0$ whenever x_{ij}^* takes the same value for all $i \in \Gamma(j)$ (see Eq. 4). Because $GMD_j \geq 0$, we must always have $f_j^{SPREAD*}(c) \geq 0$. Consequently, the solution $\{x_{ij}^* = \theta_j - c/(p_j \hat{s}_j)$ for all $i \in \Gamma(j)\}$ minimizes $f_j^{SPREAD*}(c)$ with the value of $f_j^{SPREAD*}(c) = 0$.

Part 2) To solve for the minimizer $\mathbf{x_j}$ in the definition of $f_j^{SPREAD*}(c)$, we use the fact that $y_j = d_j - \sum_{i\in\Gamma(j)} s_i x_{ij}$ to define the Lagrangian $L(\mathbf{x_j}, \zeta_j) = \sum_{i\in\Gamma(j)} \frac{V_j}{2\theta_j} s_i (x_{ij} - \theta_j)^2 + \zeta_j (p_j (d_j - \sum_{i\in\Gamma(j)} s_i x_{ij}) - c)$. From the stationarity condition $\partial L(\mathbf{x_j^*}, \zeta_j^*)/\partial x_{ij} = 0$, we have $\frac{V_j}{\theta_j} s_i (x_{ij}^* - \theta_j) - \zeta_j^* p_j s_i = 0$; thus, $x_{ij}^* = \theta_j + \zeta_j^* \frac{\theta_j p_j}{V_j}$ for all $i \in \Gamma(j)$. Substituting this x_{ij}^* into the feasibility condition $f_j^{SHORTFALL}(\mathbf{x_j}) = c$, which in this case is $p_j(d_j - \sum_{i\in\Gamma(j)} s_i x_{ij}^*) = c$, yields $\zeta_j^* = -cV_j/(d_j p_j^2)$. Since $x_{ij}^* = \theta_j + \zeta_j^* \frac{\theta_j p_j}{V_j}$ for all $i \in \Gamma(j)$, we have $x_{ij}^* - \theta_j = \zeta_j^* \frac{\theta_j p_j}{V_j} = -c/(p_j \hat{s}_j)$, and therefore $f_j^{SPREAD*}(c) = \sum_{i\in\Gamma(j)} \frac{V_j}{2\theta_j} s_i (x_{ij}^* - \theta_j)^2 = \frac{V_j}{2\theta_j} \hat{s}_j (-c/(p_j \hat{s}_j))^2 = c^2 V_j/(2 d_j p_j^2)$. Finally, $\partial f_j^{SPREAD*}(c)/\partial c = cV_j/(d_j p_j^2) > 0$ for all $c > 0$. Since $f_j^{SPREAD*}(c)$ is an increasing function of c for all $c > 0$, we have a contradiction which proves the proposition. □

A.6. Proof of Proposition 4: Existence of Split-and-Merge Invariance

Proposition 4:

1. If the parameters V_j in the baseline objective function (3a) are chosen independently of campaign demands d_j, then the optimal solution to the baseline model (SB) is not affected by arbitrary campaign splits or merges.
2. The optimal solution to the Gini model (SG) is not affected by arbitrarily splitting or merging campaigns.

Preliminaries: The following proofs use the following setup. Consider a publisher that minimizes the cost of spreading impressions over audience segments, defined as the sum of campaign-specific terms $F_j := W_j f_j$, where W_j is a campaign-specific scaling factor and f_j is either $\sum_{i\in\Gamma(j)} s_i (x_{ij} - \theta_j)^2$ for the baseline model or GMD_j for the Gini model. We have a large campaign C which we are considering splitting into two smaller campaigns A and B so that (i) campaigns A and B inherit the targeting of campaign C, i.e., $\Gamma(A) = \Gamma(B) = \Gamma(C)$, implying $\hat{s}_A = \hat{s}_B = \hat{s}_C$; and (ii) the demand of campaign C is equal to the total demand of campaigns A and B, i.e., $d_C = d_A + d_B$. Therefore, by definition $\theta_A = (d_A/d_C)\theta_C$ and $\theta_B = (d_B/d_C)\theta_C$. Furthermore, assume the allocation x_{iC} given to campaign C is proportionally split across campaigns A and B; i.e., $x_{iA} = (d_A/d_C) x_{iC}$ and $x_{iB} = (d_B/d_C) x_{iC}$, for all audience segments i. Since $x_{iA} + x_{iB} = x_{iC}$, an advertiser should be indifferent between purchasing both campaigns A and B, or just campaign C. We will now show when the objective function $\sum_j F_j$ exhibits this indifference.

Proof of Part 1 (Baseline Model): We begin with a technical lemma, then present the main proposition, and finally communicate our result using a corollary.

LEMMA EC.1. $\sum_{i\in\Gamma(C)} s_i (x_{iC} - \theta_C)^2 = \left(\frac{d_C}{d_A}\right) \sum_{i\in\Gamma(A)} s_i (x_{iA} - \theta_A)^2 + \left(\frac{d_C}{d_B}\right) \sum_{i\in\Gamma(B)} s_i (x_{iB} - \theta_B)^2$.

Proof. We simplify the left-hand-side expression as follows:

$$\sum_{i\in\Gamma(C)} s_i (x_{iC} - \theta_C)^2$$

$$
\begin{aligned}
&= \left(\frac{d_A}{d_C}\right) \sum_{i\in\Gamma(C)} s_i(x_{iC}-\theta_C)^2 + \left(\frac{d_B}{d_C}\right) \sum_{i\in\Gamma(C)} s_i(x_{iC}-\theta_C)^2, \text{ since } d_A+d_B=d_C \\
&= \left(\frac{d_A}{d_C}\right) \sum_{i\in\Gamma(A)} s_i \left(\left(\frac{d_C}{d_A}\right) x_{iA} - \left(\frac{d_C}{d_A}\right)\theta_A\right)^2 + \left(\frac{d_B}{d_C}\right) \sum_{i\in\Gamma(B)} s_i \left(\left(\frac{d_C}{d_B}\right) x_{iB} - \left(\frac{d_C}{d_B}\right)\theta_B\right)^2 \\
&= \left(\frac{d_C}{d_A}\right) \sum_{i\in\Gamma(A)} s_i \left(x_{iA}-\theta_A\right)^2 + \left(\frac{d_C}{d_B}\right) \sum_{i\in\Gamma(B)} s_i \left(x_{iB}-\theta_B\right)^2,
\end{aligned}
$$

which concludes the proof of this lemma. □

PROPOSITION EC.2. *If $W_A > W_C\left(\frac{d_C}{d_A}\right)$ and $W_B > W_C\left(\frac{d_C}{d_B}\right)$, then $F_C < F_A + F_B$; thus, an advertiser prefers to have both campaigns A and B over having only campaign C. Similarly, if $W_A < W_C\left(\frac{d_C}{d_A}\right)$ and $W_B < W_C\left(\frac{d_C}{d_B}\right)$, then $F_C > F_A + F_B$; thus, an advertiser prefers to have only campaign C over having both A and B. Moreover, if $W_A = W_C\left(\frac{d_C}{d_A}\right)$ and $W_B = W_C\left(\frac{d_C}{d_B}\right)$, then $F_C = F_A + F_B$; in this case, an advertiser is indifferent to having only C versus having both A and B.*

Proof. We prove only the first case, as all three cases are similar. Assume $W_A > W_C\left(\frac{d_C}{d_A}\right)$ and $W_B > W_C\left(\frac{d_C}{d_B}\right)$. Then:

$$
\begin{aligned}
F_C &= W_C \sum_{i\in\Gamma(C)} s_i(x_{iC}-\theta_C)^2 \\
&= W_C\left(\frac{d_C}{d_A}\right) \sum_{i\in\Gamma(A)} s_i\left(x_{iA}-\theta_A\right)^2 + W_C\left(\frac{d_C}{d_B}\right) \sum_{i\in\Gamma(B)} s_i\left(x_{iB}-\theta_B\right)^2, \text{ by Lemma EC.1} \\
&< W_A \sum_{i\in\Gamma(A)} s_i\left(x_{iA}-\theta_A\right)^2 + W_B \sum_{i\in\Gamma(B)} s_i\left(x_{iB}-\theta_B\right)^2 = F_A + F_B,
\end{aligned}
$$

which concludes the proof of this proposition. □

COROLLARY EC.1. *If $W_j = Q_j/d_j$ where the given Q_j values do not change when campaigns are split (i.e., $Q_C = Q_A = Q_B$), advertisers are indifferent toward buying small or large campaigns.*

Proof. We have $W_A d_A = Q_A = Q_C = W_C d_C$, and similarly $W_B d_B = W_C d_C$. Therefore, $W_A = W_C\left(\frac{d_C}{d_A}\right)$ and $W_B = W_C\left(\frac{d_C}{d_B}\right)$. Invoking Proposition EC.2 concludes the proof of this corollary. □

Finally, we note that as long as the weights V_j in the baseline objective (3a) are chosen independently of campaign demands d_j, then $W_j = V_j/2\theta_j$ satisfies the assumption of Corollary EC.1 with $Q_j = V_j\hat{s}_j/2$. Therefore, so long as V_j's are chosen independently of d_j's, the baseline objective is invariant to arbitrary splits and merges.

Proof of Part 2 (Gini Model): Recall from Section 3.2.2 that the spread term in our Gini objective is $\alpha\sum_j w_j G_j$, which is equivalent to $\sum_j \frac{\alpha\hat{s}_j}{2} GMD_j$. With $W_j = \frac{\alpha\hat{s}_j}{2}$, $f_j = GMD_j$, and $F_j = W_j f_j$, we aim to show that $F_C = F_A + F_B$.

Proof. Since $\hat{s}_A = \hat{s}_B = \hat{s}_C$, it is clear that $W_A = W_B = W_C$. Therefore, it remains to show that $f_C = f_A + f_B$.

$$
\begin{aligned}
f_C &= GMD_C = \frac{2}{\hat{s}_C^2} \sum_{(h,i)\in\Gamma_0(C)} s_h s_i \left|x_{hC} - x_{iC}\right| \\
&= \left(\frac{d_A}{d_C}\right) \frac{2}{\hat{s}_C^2} \sum_{(h,i)\in\Gamma_0(C)} s_h s_i \left|x_{hC} - x_{iC}\right| + \left(\frac{d_B}{d_C}\right) \frac{2}{\hat{s}_C^2} \sum_{(h,i)\in\Gamma_0(C)} s_h s_i \left|x_{hC} - x_{iC}\right| \\
&= \left(\frac{d_A}{d_C}\right) \frac{2}{\hat{s}_A^2} \sum_{(h,i)\in\Gamma_0(A)} s_h s_i \left|\left(\frac{d_C}{d_A}\right)(x_{hA} - x_{iA})\right| + \left(\frac{d_B}{d_C}\right) \frac{2}{\hat{s}_B^2} \sum_{(h,i)\in\Gamma_0(B)} s_h s_i \left|\left(\frac{d_C}{d_B}\right)(x_{hB} - x_{iB})\right| \\
&= \frac{2}{\hat{s}_A^2} \sum_{(h,i)\in\Gamma_0(A)} s_h s_i \left|x_{hA} - x_{iA}\right| + \frac{2}{\hat{s}_B^2} \sum_{(h,i)\in\Gamma_0(B)} s_h s_i \left|x_{hB} - x_{iB}\right| \\
&= GMD_A + GMD_B = f_A + f_B,
\end{aligned}
$$

which concludes the proof. □

Moreover, if the spread term $\alpha \sum_{j\in J} w_j G_j$ in our Gini objective is generalized to $\sum_{j\in J} \alpha_j w_j G_j$, then the above proposition holds so long as $\alpha_A = \alpha_B = \alpha_C$, i.e., the campaign-specific priority weighting factors α_j do not change when campaigns are arbitrarily split or merged.

A.7. Proof of Proposition 5, Part 1: Interpretation of Weights in Baseline Objective

Proposition 5, Part 1: Assuming $V_j = 1$ for all campaigns $j \in J$, the baseline objective (3a) weights campaigns by their size (i.e., impressions demanded d_j), and penalizes the average squared percentage deviation from the target allocation $(x_{ij} = \theta_j \; \forall i \in \Gamma(j))$.

Proof. Since $\frac{1}{\theta_j} = \left(\frac{1}{\theta_j}\right)^2 \cdot \theta_j = \left(\frac{1}{\theta_j}\right)^2 \cdot \frac{d_j}{\hat{s}_j}$, we have:

$$
\frac{1}{\theta_j} \sum_{i\in\Gamma(j)} s_i (x_{ij} - \theta_j)^2 = \left(\frac{1}{\theta_j}\right)^2 \cdot \frac{d_j}{\hat{s}_j} \sum_{i\in\Gamma(j)} s_i (x_{ij} - \theta_j)^2 = d_j \cdot \left(\sum_{i\in\Gamma(j)} \frac{s_i}{\hat{s}_j} \left(\frac{x_{ij}}{\theta_j} - 1\right)^2 \right).
$$

Notice that, for campaign j, the expression $\left(\frac{x_{ij}}{\theta_j} - 1\right)^2$ is the squared percentage deviation from the target allocation of audience segment i. We average this over all audience segments $i \in \Gamma(j)$ by appropriately weighting larger segments more than smaller segments (note that $\sum_{i\in\Gamma(j)} \frac{s_i}{\hat{s}_j} = 1$). □

Appendix B: Additional Material: Background, Plots, Descriptions, and Extended Analysis

B.1. Bar Graph Visualization of Impression Spread

Consider an advertiser who wants to visualize how well-spread the impressions she got are over some dimension of interest, for example, geography. Her ad campaign targeted only the three western states and received a total of 10 million impressions, with the breakdown being 7, 1, and 2 million from California, Oregon, and Washington, respectively. Although we can visualize this data using a bar graph, as it could be done by using the results of the baseline model, this is problematic for at least two reasons. First, the geographic dimension does not have a natural order, and thus, there are six different permutations of the bars (see Figure EC.1), all of which are equally valid. Second, a typical bar graph is not normalized for differing population sizes across audience segments. Indeed, in our example California, Oregon, and Washington have populations of 39, 4, and 7 million people, respectively; this corresponds to 78%, 8% and 14% of the western region's population of 50 million. An evenly-spread ad campaign of 10 million impressions would assign 7.8, 0.8, and 1.4 million impressions, respectively, to California, Oregon, and Washington. Meanwhile, our original allocation of 7, 1, and 2 million impressions is 10.3% below, 25% above, and 42.9% above the equal-proportion solution. In essence, the bar graph is misleading since it incorrectly compares the 7 million impressions in California with the 2 million in Washington, without accounting for the fact that California and Washington's audience sizes are on different scales (indeed – California's allocation is in fact below average, not above average!). A Lorenz curve, on the other hand, visually represents the spread of a categorical distribution in such a way that, unlike the more widely used (and abused) bar graph, does not suffer from (a) multiple possible visual representations due to arbitrary ordering of categories, and (b) poor scaling.

Note that the practical importance of visualizing spread using a Lorenz curve instead of a bar graph is compounded when comparing spread across multiple ad campaigns. This is because while it may be possible for users to agree on a particular ordering of categories (e.g., listing all geographic regions alphabetically by name from left-to-right), different campaigns target different subsets of categories along a dimension. For example, in a bar graph, the third bar from the left could represent California for one campaign, and Kansas for another. Indeed, a bar graph would only properly represent spread of impressions when (a) all campaigns target all audience segments, (b) users can agree on a single arbitrary ordering of the segments within a bar graph, and (c) all audience segments represent audiences of the same size. Clearly, this special case is quite rare in practice, which motivates the Lorenz curve as a means of visually representing the spread of impressions across audience segments.

B.2. Alternative Models

In this section, we discuss several alternative spreading objectives, and show how seemingly small structural modifications affect the ability of the corresponding model to satisfy the five key properties listed in Table 1.

Campaign-specific scaling factors, i.e., the w_j's in the Gini objective $\alpha \sum_{j \in J} w_j G_j$ and W_j's in the baseline objective $\sum_{(i,j) \in \Gamma} W_j s_i (x_{ij} - \theta_j)^2$, must be chosen appropriately for several key properties to hold. Consider the following model variants with modified scaling factors, and the resulting effects:

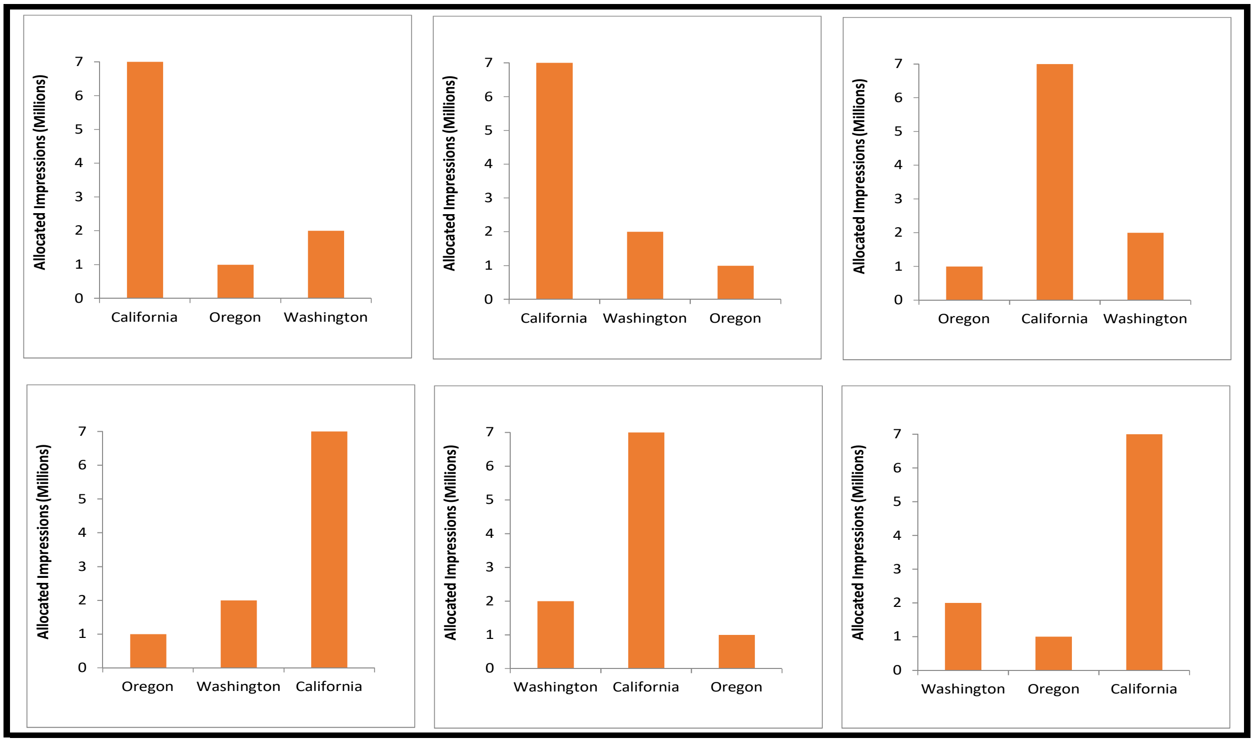

Figure EC.1 Impression spread by geographic region, visually represented using bar graphs.

- **Alternative Gini Objective,** $\alpha \sum_{j\in J} d_j G_j$: If we weight the Gini coefficients by the demands d_j instead of by impressions allocated $w_j = d_j - y_j$, the Gini term in the objective would be equivalent to $\alpha \sum_{j\in J}(y_j + w_j)G_j$. Consequently, demand shortfalls y_j would be penalized by the spread metric αG_j, violating Property 3 (Sufficient Orthogonality). Moreover, whereas $\sum_{j\in J} w_j G_j$ is linearizeable via the formulation (7)-(9), the expression $\sum_{j\in J} d_j G_j$ is non-convex in the primary decision variables x_{ij}, causing this model variant to additionally violate Property 1 (Efficient Solvability).
- **Alternative Baseline Objective,** $\sum_{(i,j)\in\Gamma} V_j\theta_j s_i (x_{ij} - \theta_j)^2$: In this model variant, the scaling factor is $W_j = V_j\theta_j$ rather than $W_j = V_j/(2\theta_j)$, and as before we assume V_j is independent of demand d_j. This objective has a nice interpretation, since with $V_j = 1$ it simplifies to $\sum_{(i,j)\in\Gamma} d_j (s_i/\hat{s}_j)(x_{ij} - \theta_j)^2$, which is a weighted average of the squared deviations $(x_{ij} - \theta_j)$, computed using the audience segment sizes s_i as weights, and then scaling by the size of the campaign d_j (recall that by definition, $\sum_{i\in\Gamma(j)}(s_i/\hat{s}_j) = 1$). Unfortunately, this objective violates Property 4 (Split-and-Merge Invariance), which can be verified by tracing the proof of Proposition 4. In this case, merged campaigns receive improved performance, and to avoid this distortion the publisher would need to refrain from merging campaigns together to make the planning problem more efficient to solve.
- **Alternative Baseline Objective,** $\sum_{(i,j)\in\Gamma} V_j/(d_j\theta_j) s_i (x_{ij} - \theta_j)^2$: In this model variant, the scaling factor is $W_j = V_j/(d_j\theta_j)$ rather than $W_j = V_j/(2\theta_j)$, and as before we assume V_j is independent of demand d_j. Again, Property 4 (Split-and-Merge Invariance) is violated. But this time, split campaigns receive improved performance. In Appendix B.3 we show that the incentive that advertisers have to split a large campaign into

many smaller ones can be significant. Indeed, splitting one campaign into 10 smaller copies can yield 12% lower spread cost and 69% lower shortfall, while splitting it into 100 smaller copies yields 87% lower spread cost and 96% lower shortfall, all at the expense of other campaigns.

Other spreading objectives could also be considered, but many have similar issues. For example:

- **Alternative Baseline Objective,** $\sum_{(i,j)\in\Gamma} V_j/(2\mu_j)s_i(x_{ij}-\mu_j)^2$: This is the same as our original baseline objective, except we have replaced the nominal target θ_j (a constant) with the mean allocation $\mu_j = (1/\hat{s}_j)\sum_{i\in\Gamma(j)} x_{ij}$ (a dependent variable). Unlike the original, because this objective measures spread as a distance from the mean allocation rather than an a priori-defined target, it satisfies both Properties 2 and 3 (Ideal Allocation when Possible and Sufficient Orthogonality). However, because the associated model is no longer a quadratic program, Property 1 (Efficient Solvability) is violated. Although the model is technically still convex, the $(1/\mu_j)$ factors lead to some numerical difficulties, especially when one or more μ_j's are small. On the other hand, the $(1/\mu_j)$ factors are required for Property 4 (Split-and-Merge Invariance). Consequently, there does not exist a quadratic spreading objective that simultaneously (i) penalizes deviations from μ_j, (ii) is efficiently solvable, and (iii) is split-and-merge invariant.

Finally, we mention two spread objectives that involve a sum of absolute values like our Gini objective but are structurally more similar to our baseline, in that they minimize absolute rather than squared deviations:

- **Alternative Absolute Deviation Objective,** $\sum_{(i,j)\in\Gamma} V_j s_i|x_{ij}-\theta_j|$: Because this objective, like our baseline, measures spread as a distance from an a priori-defined target, it violates Properties 2 and 3 (Ideal Allocation when Possible and Sufficient Orthogonality). However, the other three properties are satisfied. The corresponding model is representable as a linear program, and therefore efficiently solvable. The scaling factor V_j satisfies split-and-merge invariance, and although not immediately obvious, a rearrangement like that in the proof of Proposition 5 part 1 shows that this objective can be interpreted as weighting each campaign by its size, d_j.
- **Alternative Absolute Deviation Objective,** $\sum_{(i,j)\in\Gamma} V_j s_i|x_{ij}-\mu_j|$: Because this objective measures spread as a distance from the mean allocation, it satisfies Properties 2 and 3 (Ideal Allocation when Possible and Sufficient Orthogonality). Moreover, the other three properties are satisfied as well.

Note that, in the existing literature, there are no single-period ad allocation models that we are aware of which spread impressions using either of the two objectives above. Some multi-period models minimize expressions similar to $|x_{ij}-\theta_j|$, and for this reason we use such a baseline in our multi-period extension in Appendix E. We are not aware of any ad planning models that have previously explored the use of the second objective, which minimizes absolute distances to mean allocations, μ_j. We leave a full analysis of this new model to future work, and focus here on comparing our Gini model to the most commonly-used single-period baseline, (SB), as well as deriving computational results which are specific to deploying the Gini model (i.e., our decomposition method and related structural results). One particular point of departure between the absolute deviation models above and our Gini model that could be important in practice is that the absolute deviation models do not produce unique solutions (it is easy to construct examples with multiple optimal solutions). On the other hand, we conjecture that our Gini model produces unique solutions (as it has in all examples that we tested).

B.3. Quantifying the Importance of Split-and-Merge Invariance

We now introduce an alternative baseline model to estimate the inefficiency associated with a model that is not merge-and-split invariant (recall that both the baseline and Gini models are merge-and-split invariant). The objective function, which induces an incentive for advertisers to split their campaigns to get better performance, is:

$$\min \sum_{(i,j)\in\Gamma} V_j/(d_j\theta_j)s_i(x_{ij}-\theta_j)^2 + \sum_{j\in J} p_j y_j \tag{EC.4}$$

Note that in this model variant, the scaling factor W_j in the spread term $\sum_{(i,j)\in\Gamma} W_j s_i(x_{ij}-\theta_j)^2$ of the baseline objective has been set to $W_j = V_j/(d_j\theta_j)$ rather than $W_j = V_j/(2\theta_j)$. Furthermore, we fix V_j to be equal to 16887 (the average campaign demand $(1/|J|)\sum_{j'\in J} d_{j'}$ for all campaigns, $j \in J$, of the original non-split instance), which produces solutions that balance both shortfall and spread.

Using this alternative baseline model, we first solve a simulated instance with 20 campaigns and 100 viewer types (this is one of the locally tight instances described in Section 5.1). Then, we split campaign #1 into 10 campaigns that each have 1/10th the demand of the original, and re-solve. Finally, we split campaign #1 into 100 campaigns that each have 1/100th the demand of the original, and re-solve. We then look at the combined contribution from all copies of campaign #1 to the alternative baseline's optimal value, and compare this to the contribution from all other campaigns. The results are in Table EC.3 below.

Optimal Value of the Alternative Baseline Objective			
	Original (No Splitting)	After 10 Splits	After 100 Splits
Contribution of All Copies of Campaign #1	102.3 (21.5%)	44.7 (8.3%)	6.0 (1.1%)
Contribution of All Other Campaigns	372.9 (78.5%)	493.2 (91.7%)	554.4 (98.9%)
Total	475.3 (100%)	537.9 (100%)	560.4 (100%)

Table EC.3 Contribution of campaign #1 to the alternative baseline's optimal value. Numerical values are optimal values of (SB) using the alternative objective function (EC.4).

Recall that the baseline objective minimizes spread cost and shortfall cost. Therefore, when campaign #1 contributes less to the optimal value, it means that it receives better spread and/or lower shortfall. We can also repeat the above analysis, isolating the effect on the spread cost and shortfall cost components of the objective. Doing so, we find that both spreading and shortfall are improved for campaign #1, at the expense of the other campaigns, and the magnitude of this improvement increases as a function of the number of times that campaign #1 is split. Details are in Tables EC.4 and EC.5 below.

If advertisers are fully rational, one can imagine that they could split each campaign into tens or hundreds of thousands of tiny campaigns to get even better service from a publisher that uses an objective function that is not invariant to arbitrary splits (i.e., if they wrote code on their end to perform this splitting, and then recombine the

Spread Cost Component of the Alternative Baseline Objective			
	Original (No Splitting)	After 10 Splits	After 100 Splits
Contribution of All Copies of Campaign #1	23.2 (8.6%)	20.5 (5.8%)	3.0 (0.8%)
Contribution of All Other Campaigns	247.3 (91.4%)	333.8 (94.2%)	371.0 (99.2%)
Total	270.5 (100%)	354.3 (100%)	374.0 (100%)

Table EC.4 Contribution of campaign #1 to the spread cost component of the alternative baseline's optimal value. Numerical values are spread costs as defined in the first term of the alternative objective function (EC.4).

Shortfall Cost Component of the Alternative Baseline Objective			
	Original (No Splitting)	After 10 Splits	After 100 Splits
Contribution of All Copies of Campaign #1	79.1 (38.6%)	24.2 (13.2%)	2.9 (1.6%)
Contribution of All Other Campaigns	125.6 (61.4%)	159.4 (86.8%)	183.4 (98.4%)
Total	204.8 (100%)	183.6 (100%)	186.4 (100%)

Table EC.5 Contribution of campaign #1 to the shortfall cost component of the alternative baseline's optimal value. Numerical values are shortfall costs as defined in the second term of the alternative objective function (EC.4).

results for reporting). This is why we suggest that publishers should use objectives that are invariant to arbitrary splits and merges.

B.4. Detailed Description of Residual Distribution Comparison

In this section, we describe the residuals of the baseline and Gini models in detail, and additionally include plots for the Globally Tight and Loose instances, which were omitted from the main body of the paper to save space.

We begin with an in-depth description of Figure 6, which illustrates, for the family of Locally Tight instances, how the residual distributions of the Gini and baseline models change as a function of revenue ($\%R$). From the top-left of Figure 6, we see that although the Gini model (blue) produces solutions that correspond to all revenue levels $\%R \in [0.8, 1]$, the baseline model (red) only produces solutions in the range $\%R \in [0.868, 1]$. The baseline model's revenue range is more limited than the Gini's because the baseline objective is not sufficiently orthogonal (c.f., Property 3). Indeed, starting from $\nu = \epsilon$, where ϵ is a very small positive quantity, yields minimal shortfall and consequently maximal revenue ($\%R = 1$) for the baseline model. Increasing ν exerts more effort on spreading and so, up to a point, shortfall increases and consequently revenue decreases. However, after some point it is not possible to get better spreading by reducing revenue further, and as ν grows large we converge to a solution with $\%R = 0.868$. On the other hand, the Gini objective is separable in shortfall and spread; thus, it is always possible to achieve better spreading by increasing shortfall. For the Gini model, as we increase α above ϵ we can always reach a point where all residuals shrink down to zero, yielding perfect spreading. This point may be very low-revenue, but it is always attainable, and often such a point exists well before $\%R$ drops to zero (for the loose family of instances, the Gini model attains perfect spreading at approximately $\%R = 0.85$).

Continuing to describe the solution structure in the locally tight case, we observe that in the top-left of Figure 6 the top-most blue line (95th-percentile of Gini) is above the top-most red line (95th percentile of baseline).

As well, the bottom-most blue line (5th percentile of Gini) is below the bottom-most red line (5th percentile of baseline). This relationship holds for all revenue levels over which both models provide solutions ($\%R \in [0.868, 1]$). Thus, for all revenue levels, the residual distribution is wider under the Gini model than under the baseline model. That said, although the Gini's distribution has wider tails, it also has more mass in its midsection. Compare the third pair of lines from the bottom (25th percentiles, dotted) and notice that the blue Gini line is above the red baseline, with the Gini line approaching the zero-level faster than the baseline as revenue decreases. The second pair of lines from the top (90th percentiles, dashed) exhibit the same pattern, with the blue Gini line below (i.e., closer to the zero-level) than the baseline. The midsection of the distribution is quite concentrated, as can be seen more clearly by zooming in on the portion of the plot where $\%R \in [0.95, 1]$ and residuals are in the range $[-0.1, 0.1]$, presented in the bottom-left panel of Figure 6. Notice that the 40th percentile of Gini (blue, dashed) starts (at $\%R = 1$) just above the 40th percentile of baseline (red, dashed), at a value of -0.0669. As revenue decreases, we see that the 40th percentile of Gini converges to the zero-level at $\%R = 0.95$, whereas the 40th percentile of baseline actually drops a little from -0.0729 at $\%R = 1$ to -0.0749 at $\%R = 0.95$. Notably, the median (50th percentile) of Gini (thin solid blue line) is always at the zero-level for all revenue levels, whereas the median of baseline (thin solid red line) is always slightly below the zero-level. The 60th percentile of Gini (dashed blue line) is also horizontal, at the zero-level for all revenue levels, compared to the 60th percentile of baseline (dashed red line) which runs from 0.0110 at $\%R = 1$ to 0.0090 at $\%R = 0.95$.

Another way to compare the residual distributions is by overlaying histograms of their density functions. Continuing to describe the locally tight instance, we observe that the top-right of Figure 6 has three such overlayed histograms, with blue (Gini) histograms overlayed onto red (baseline) histograms, and areas of agreement showing in purple. The first overlayed histogram compares the residual distributions at $\%R = 1$, and corresponds to the cross-section in the top-left plot indicated by the right-most vertical grey bar. From the tallest bar at the center of the histogram, we can see that 24.1% of Gini's residuals are in the range $[-0.035, 0.035]$, compared to 16.4% from the baseline model. The second overlayed histogram compares the residual distributions at $\%R = 0.95$, and corresponds to the cross-section in the top-left plot indicated by the middle vertical grey bar. Here, we can see that 38.2% of Gini's residuals are in the range $[-0.035, 0.035]$, compared to 21.6% from the baseline model. Finally, the third overlayed histogram compares the residual distributions at $\%R = 0.9$, and corresponds to the cross-section in the top-left plot indicated by the left-most grey bar. Here, we can see that 51.9% of Gini's residuals are in the range $[-0.035, 0.035]$, compared to 23.9% from the baseline model.

In summary, the Gini's residual distribution is significantly more concentrated in the middle near zero, and slightly fatter in the tails than the baseline. This difference is the most striking for the locally tight family of instances; however, this structural result is robust and holds for the globally tight and loose families of instances as well (see Table 3). For completeness, Figure EC.2 plots the residual distributions for the globally tight and loose families of instances.

Finally, in Section 5.5 we investigated the residual distribution under the case of having an outside option that we can also sell impressions to. Figure 7 plotted the residual distribution for the locally tight family of instances.

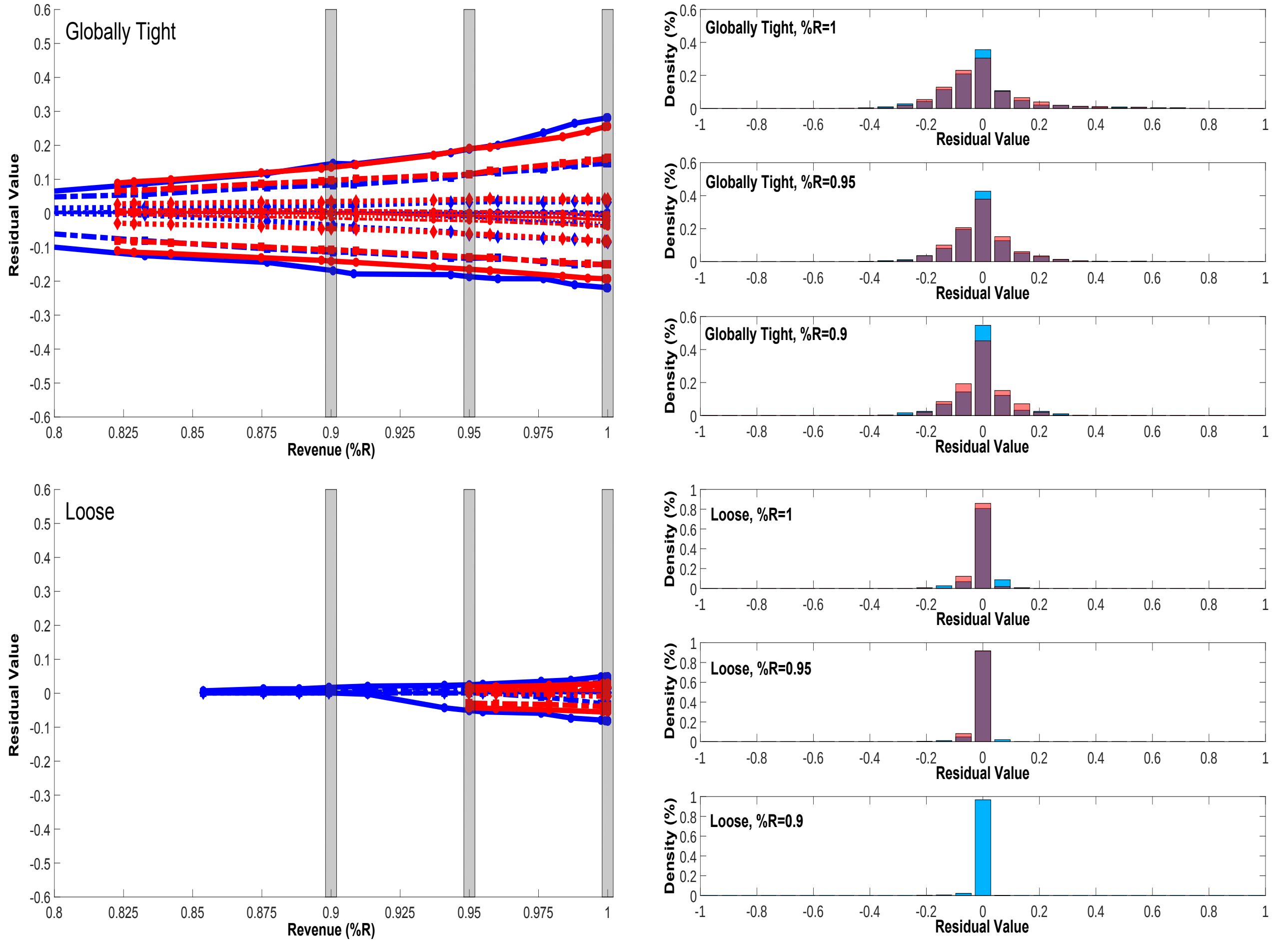

Figure EC.2 Residual Distribution Comparison (Globally Tight and Loose Instances). **Top Row**: Results for the globally tight instances. **Bottom Row**: Results for the loose instances. **Left Column**: Percentiles of the residual distributions from the baseline (red) and Gini (blue) models, as a function of revenue ($\%R$); refer to the legend in Figure 6 of §5.4, which represents each percentile by a different dashed pattern. **Right Column**: Histograms depict residual distributions at $\%R = 1$, 0.95, and 0.9; each histogram corresponds to a cross-section in the left column highlighted by a vertical grey bar. Histograms for the Gini (blue) and baseline (red) distributions are overlayed, with purple bars indicating where the blue and red bars overlap.

To save space in the main body of the manuscript, we omitted the globally tight and loose cases; these are plotted in Figure EC.3.

B.5. The Supply-Relaxed Subproblem as a Market-Based Model

Small to medium ad aggregators, who do not operate their own websites but instead purchase all of their impression traffic from other publishers and sell it to advertisers, may also use our Gini-based model (SG) to plan and allocate impressions to advertisers who have purchased guaranteed contracts and expect to receive well-spread impressions. In this context, we assume the aggregator has estimates for the market prices $\hat{\beta}_i$ and market supplies s_i of each audience segment. In addition, we assume the volume of impression traffic the aggregator handles is small relative to the size of the market, so that all supply constraints (3c) are nonbinding. The resulting

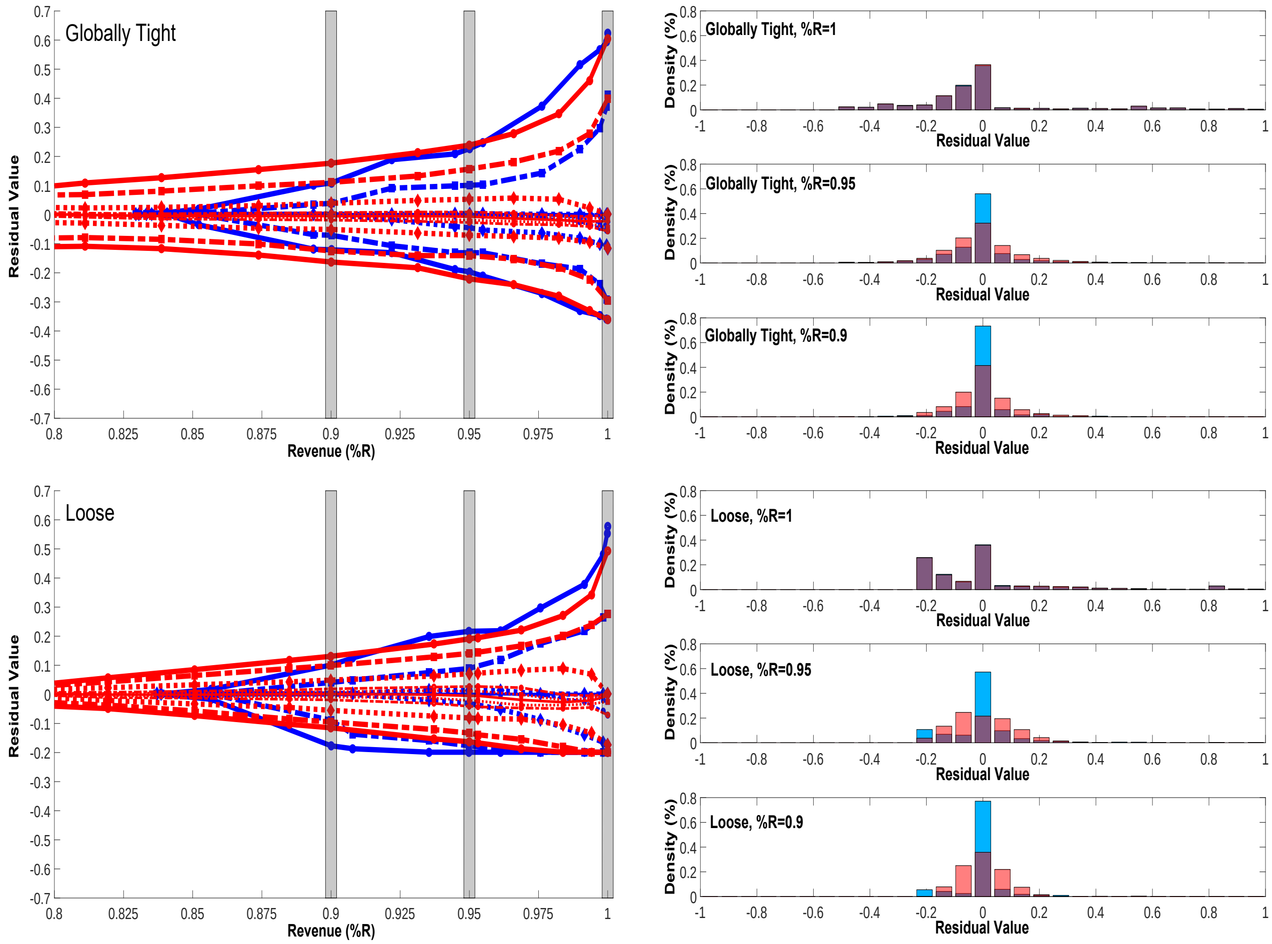

Figure EC.3 Residual Distribution Comparison with Outside Option (Globally Tight and Loose Instances). **Top Row**: Results for the globally tight instances. **Bottom Row**: Results for the loose instances. **Left Column**: Percentiles of the residual distributions from the baseline (red) and Gini (blue) models, as a function of revenue ($\%R$); refer to the legend in Figure 6 of §5.4, which represents each percentile by a different dashed pattern. **Right Column**: Histograms depict residual distributions at $\%R = 1$, 0.95, and 0.9; each histogram corresponds to a cross-section in the left column highlighted by a vertical grey bar. Histograms for the Gini (blue) and baseline (red) distributions are overlayed, with purple bars indicating where the blue and red bars overlap.

Gini-based model is precisely (SG) with the objective $\min \alpha \sum_{j\in J} \frac{1}{\hat{s}_j} \sum_{(h,i)\in\Gamma_0(j)} s_h s_i |x_{hj} - x_{ij}| + \sum_{j\in J} p_j y_j + \sum_{(i,j)\in\Gamma} \hat{\beta}_i s_i x_{ij}$ and the supply constraints dropped.

It is important to note that this market-based model's optimal solution can be obtained by solving (PS) with $\psi_j = \alpha/\hat{s}_j$. Indeed, (PS) is exactly this market-based model with additional constant terms $-\sum_{i\in I} \hat{\beta}_i s_i$ in the objective, and the presence of these constant terms does not affect the optimal solution. Moreover, since (PS) decomposes by campaign, Theorem 1 tells us the optimal allocation for each campaign j in this market-based model. Indeed, since in this case prices $\hat{\beta}_i$ are exogenously given and supply constraints are nonbinding, there is no need to iterate through Dantzig-Wolfe decomposition, and the solution is immediate. Using the optimal solution x^*_{ij}, the aggregator can then establish impression targets $w^*_{ij} = s_i x^*_{ij}$ for each audience segment and campaign pair, which it would attempt to achieve by acquiring impressions at or near predicted market prices.

We believe this model provides a salient first-order characterization of an aggregator's optimal policy, and that suitable data exists to estimate the model's parameters. Although market prices may fluctuate over time, we expect that historical data could be used to estimate the prices that other market participants will charge in expectation for impressions from popular audience segments. As well, estimates of audience sizes can be obtained from third-party audience measurement firms such as ComScore or Nielsen. Note that since there are no supply constraints in this variant of our model, audience size estimates are used exclusively in the objective function to measure the quality of spread.

There are of course many nuanced complications concerning intermediation in online advertising. For more details, see Balseiro et al. (2015b) and Allouah et al. (2017).

References

Allouah A, Besbes O (2017) Auctions in the online display advertising chain: A case for independent campaign management (June 5, 2017). Working Paper, Columbia University. Available at SSRN: `https://papers.ssrn.com/sol3/papers.cfm?abstract\_id=2919665`.

Balseiro S, Candogan O, Gurkan H (2015) Multi-stage intermediation in online internet advertising (September 21, 2018). Working Paper, Columbia University. Available at SSRN: `http://dx.doi.org/10.2139/ssrn.2661459`.

Appendix C: Decomposition Method

We now derive in detail our decomposition scheme for our Gini ad allocation problem (SG). We begin by restating the general formulation of this single-period problem, which we presented in §6:

$$
\begin{aligned}
\text{(P-ORIG)} \qquad z^* = \min & \sum_{j\in J,(h,i)\in\Gamma_0(j)} \psi_j s_h s_i |x_{hj} - x_{ij}| + \sum_j p_j y_j \\
\text{s.t.} & \sum_{i\in\Gamma(j)} s_i x_{ij} + y_j = d_j \qquad \forall j \in J \qquad \text{(demand)} \\
& \sum_{j\in\Gamma(i)} s_i x_{ij} \le s_i \qquad \forall i \in I \qquad \text{(supply)} \\
& x_{ij} \ge 0\,\forall (i,j)\in\Gamma; \qquad y_j \ge 0\,\forall j \in J
\end{aligned}
$$

First, in Section C.1, we describe how to disaggregate the audience segments of (P-ORIG) into agglomerations of individual impressions. Then, in the subsequent subsection, we use this impression-based formulation to formally derive our decomposition method.

C.1. Translation to Disaggregate Impression-Level Space

Because some of our proofs are clearer when the quantities of interest are impressions rather than buckets of impressions aggregated into audience segments, we first translate (P-ORIG) into an equivalent impression-level optimization problem. Let R be the set of all impressions, R_i be the set of impressions that comprise audience segment i, Λ be the set of (impression, campaign) pairs that define the targeted instance, $\Lambda(r)$ be the set of campaigns that are targeted by impression r, $\Lambda(j)$ be the set of impressions targeted by campaign j, and $\Lambda_0(j) = \{(q,r)\in\Lambda(j)^2 : q < r\}$ index all pairs of impressions targeted by campaign j. Note that by definition, (i) $|R_i| = s_i$, (ii) $|\Lambda(j)| = \hat{s}_j$, and (iii) $\Lambda(r) = \Gamma(i)$ when impression r is in audience segment i. Our impression-level formulation is as follows (Table EC.2 provides a quick reference to all impression-level notation):

$$
\begin{aligned}
\text{(P2)} \qquad \min & \sum_{j\in J,(q,r)\in\Lambda_0(j)} \psi_j |x_{qj} - x_{rj}| + \sum_j p_j y_j \\
\text{s.t.} & \sum_{r\in\Lambda(j)} x_{rj} + y_j = d_j \qquad \forall j \in J \qquad \text{(demand)} \\
& \sum_{j\in\Lambda(r)} x_{rj} \le 1 \qquad \forall r \in R \qquad \text{(supply)} \\
& x_{rj} \ge 0\,\forall (r,j)\in\Lambda; \qquad y_j \ge 0\,\forall j \in J
\end{aligned}
$$

The following proposition formalizes the connection between the audience-level formulation (P-ORIG) and the impression-level formulation (P2).

PROPOSITION EC.3. *Problem (P2) with (i) additional variables $x_{ij}, (i,j)\in\Gamma$, and (ii) additional constraints $x_{rj} = x_{ij}\;\forall r\in R_i, \forall j\in J$, which force all impression allocations within the same audience segment to have the same value, is equivalent to (P-ORIG).*

Proof. We have previously shown (c.f., Appendix A.3) that the objective functions of (P2) and (P-ORIG) are equivalent when $x_{rj} = x_{ij}\ \forall r \in R_i, j \in J$. In a similar manner, the left-hand-side of (P2)'s demand constraint aggregates into an audience-level expression as follows:

$$\sum_{r\in\Lambda(j)} x_{rj} = \sum_{i\in\Gamma(j)}\sum_{r\in R_i} x_{rj} = \sum_{i\in\Gamma(j)}\sum_{r\in R_i} x_{ij} = \sum_{i\in\Gamma(j)} |R_i| x_{ij} = \sum_{i\in\Gamma(j)} s_i x_{ij},$$

and the left-hand-side of (P2)'s supply constraint aggregates into an audience-level expression as follows:

$$\sum_{j\in\Lambda(r)} x_{rj} = \sum_{j\in\Lambda(r)} x_{ij} = \sum_{j\in\Gamma(i)} x_{ij}.$$

Since the objective function and all constraints are logically equivalent, the result follows. □

Technically, (P2) is a relaxation of (P-ORIG), and is equivalent to (P-ORIG) only when we additionally impose the constraints $x_{rj} = x_{ij}\ \forall r \in R_i, j \in J$. However, it turns out that we can impose these constraints in (P2) without loss of optimality. That is, there always exists an optimal solution to (P2) where, for each campaign $j \in J$, the allocation x_{rj} is equal across all impressions $r \in R_i$ within each audience segment $i \in I$. Consequently, (P2) and (P-ORIG) can be treated as essentially equivalent. To prove this result, we first prove the following technical lemma, which shows that if an unequal solution is made more equal, then the value of the Gini term in the objective function of (P2) improves (reduces). The proof requires several steps to be rigorously shown, and formalizes how we may incrementally improve a solution by making it more equal.

LEMMA EC.2. *Assume we are given an impression-level allocation vector $\mathbf{x}_j \equiv \{x_{rj}\}_{r\in\Lambda(j)}$ for some campaign $j \in J$ whose components are not all equal, i.e., there exists a pair of impressions (q, q') such that $x_{qj} < x_{q'j}$. Let $\Delta \in (0, \frac{1}{2}(x_{q'j} - x_{qj})]$ and construct an alternative impression-level allocation vector $\overline{\mathbf{x}}_j \equiv \{\overline{x}_{rj}\}_{r\in\Lambda(j)}$ such that $\overline{x}_{q'j} = x_{q'j} - \Delta$, $\overline{x}_{qj} = x_{qj} + \Delta$, and $\overline{x}_{rj} = x_{rj}\ \forall r \notin \{q, q'\}$. Then $\sum_{(r,r')\in\Lambda_0(j)} |\overline{x}_{rj} - \overline{x}_{r'j}| < \sum_{(r,r')\in\Lambda_0(j)} |x_{rj} - x_{r'j}|$, i.e., the GMD value of $\overline{\mathbf{x}}_j$ is strictly lower (better) than the GMD value of $\mathbf{x}_j$.*

Proof. Without loss of generality, assume the impressions are ordered by increasing value of x_{rj}; that is, $x_{1j} \le x_{2j} \le \cdots \le x_{\hat{s}_j j}$. Partition the impressions $r \in \Lambda(j) \setminus \{q, q'\}$ into five sets:

$$A = \{r \in \Lambda(j) : x_{rj} \le x_{qj}, r \ne q\},$$
$$B = \{r \in \Lambda(j) : x_{qj} < x_{rj} \le \overline{x}_{qj}\},$$
$$C = \{r \in \Lambda(j) : \overline{x}_{qj} < x_{rj} \le x_{q'j}\},$$
$$D = \{r \in \Lambda(j) : \overline{x}_{q'j} < x_{rj} \le x_{q'j}, r \ne q'\}, \text{ and}$$
$$E = \{r \in \Lambda(j) : x_{q'j} < x_{rj}\}.$$

By construction, these sets are mutually exclusive, none of them contain the impressions q or q', and $A \cup B \cup C \cup D \cup E \cup \{q, q'\} = \{1, \ldots, \hat{s}_j\} = \Lambda(j)$. Moreover, impressions are ordered according to $A \prec q \prec B \prec C \prec$

$D \prec q' \prec E$, where $A \prec B$ means that all impressions in A precede all impressions in B with respect to the established ordering. Using these definitions, we can simplify the GMD value of $\overline{\mathbf{x}}_j$ as follows:

$$\begin{aligned}
\sum_{(r,r')\in\Lambda_0(j)} |\overline{x}_{rj}-\overline{x}_{r'j}| = & \sum_{(r,r')\in\Lambda_0(j):r\neq q,r'\neq q'} |\overline{x}_{rj}-\overline{x}_{r'j}| + \sum_{r\in A,r'=q} |\overline{x}_{rj}-\overline{x}_{r'j}| \\
& + \sum_{r=q,r'\in B} |\overline{x}_{rj}-\overline{x}_{r'j}| + \sum_{r=q,r'\in C\cup D\cup E} |\overline{x}_{rj}-\overline{x}_{r'j}| + \sum_{r\in A\cup B\cup C,r'=q'} |\overline{x}_{rj}-\overline{x}_{r'j}| \\
& + \sum_{r\in D,r'=q'} |\overline{x}_{rj}-\overline{x}_{r'j}| + \sum_{r=q',r'\in E} |\overline{x}_{rj}-\overline{x}_{r'j}| + \sum_{r=q,r'=q'} |\overline{x}_{rj}-\overline{x}_{r'j}| \\
= & \sum_{(r,r')\in\Lambda_0(j):r\neq q,r'\neq q'} |x_{rj}-x_{r'j}| + \sum_{r\in A} |x_{rj}-\overline{x}_{qj}| \\
& + \sum_{r'\in B} |\overline{x}_{qj}-x_{r'j}| + \sum_{r'\in C\cup D\cup E} |\overline{x}_{qj}-x_{r'j}| + \sum_{r\in A\cup B\cup C} |x_{rj}-\overline{x}_{q'j}| \\
& + \sum_{r\in D} |x_{rj}-\overline{x}_{q'j}| + \sum_{r'\in E} |\overline{x}_{q'j}-x_{r'j}| + |\overline{x}_{qj}-\overline{x}_{q'j}| \\
= & \sum_{(r,r')\in\Lambda_0(j):r\neq q,r'\neq q'} |x_{rj}-x_{r'j}| + \sum_{r\in A} (\overline{x}_{qj}-x_{rj}) \\
& + \sum_{r'\in B} (\overline{x}_{qj}-x_{r'j}) + \sum_{r'\in C\cup D\cup E} (x_{r'j}-\overline{x}_{qj}) + \sum_{r\in A\cup B\cup C} (\overline{x}_{q'j}-x_{rj}) \\
& + \sum_{r\in D} (x_{rj}-\overline{x}_{q'j}) + \sum_{r'\in E} (x_{r'j}-\overline{x}_{q'j}) + (\overline{x}_{q'j}-\overline{x}_{qj}) \\
= & \sum_{(r,r')\in\Lambda_0(j):r\neq q,r'\neq q'} |x_{rj}-x_{r'j}| + \sum_{r\in A} (x_{qj}+\Delta-x_{rj}) \\
& + \sum_{r'\in B} (x_{qj}+\Delta-x_{r'j}) + \sum_{r'\in C\cup D\cup E} (x_{r'j}-x_{qj}-\Delta) + \sum_{r\in A\cup B\cup C} (x_{q'j}-\Delta-x_{rj}) \\
& + \sum_{r\in D} (x_{rj}-x_{q'j}+\Delta) + \sum_{r'\in E} (x_{r'j}-x_{q'j}+\Delta) + ((x_{q'j}-\Delta)-(x_{qj}+\Delta)) \\
= & \sum_{(r,r')\in\Lambda_0(j):r\neq q,r'\neq q'} |x_{rj}-x_{r'j}| + \sum_{r\in A} (x_{qj}-x_{rj}) \\
& + \sum_{r'\in B} (x_{qj}-x_{r'j}) + \sum_{r'\in C\cup D\cup E} (x_{r'j}-x_{qj}) + \sum_{r\in A\cup B\cup C} (x_{q'j}-x_{rj}) \\
& + \sum_{r\in D} (x_{rj}-x_{q'j}) + \sum_{r'\in E} (x_{r'j}-x_{q'j}) + (x_{q'j}-x_{qj}) - 2\Delta(|C|+1) \\
= & \sum_{(r,r')\in\Lambda_0(j):r\neq q,r'\neq q'} |x_{rj}-x_{r'j}| + \sum_{r\in A} |x_{qj}-x_{rj}| \\
& - \sum_{r'\in B} |x_{qj}-x_{r'j}| + \sum_{r'\in C\cup D\cup E} |x_{r'j}-x_{qj}| + \sum_{r\in A\cup B\cup C} |x_{q'j}-x_{rj}| \\
& - \sum_{r\in D} |x_{rj}-x_{q'j}| + \sum_{r'\in E} |x_{r'j}-x_{q'j}| + |x_{q'j}-x_{qj}| - 2\Delta(|C|+1) \\
< & \sum_{(r,r')\in\Lambda_0(j):r\neq q,r'\neq q'} |x_{rj}-x_{r'j}| + \sum_{r\in A} |x_{qj}-x_{rj}| \\
& + \sum_{r'\in B} |x_{qj}-x_{r'j}| + \sum_{r'\in C\cup D\cup E} |x_{r'j}-x_{qj}| + \sum_{r\in A\cup B\cup C} |x_{q'j}-x_{rj}| \\
& + \sum_{r\in D} |x_{rj}-x_{q'j}| + \sum_{r'\in E} |x_{r'j}-x_{q'j}| + |x_{q'j}-x_{qj}|
\end{aligned}$$

$$= \sum_{(r,r')\in\Lambda_0(j)} |x_{rj} - x_{r'j}|,$$

which is the GMD value of $\mathbf{x}_j$. The first equality expands $\sum_{(r,r')\in\Lambda_0(j)} |\overline{x}_{rj} - \overline{x}_{r'j}|$ into eight terms, allowing us to treat pairs of impressions (r, r') differently based on their membership in the sets $\{A, B, C, D, E\}$. The second equality follows since $\overline{x}_{rj} = x_{rj}$ for all $r \in A \cup B \cup C \cup D \cup E$. The third equality follows from the definitions of the sets $\{A, B, C, D, E\}$, and the fourth is by substitution. The fifth equality follows by collecting all Δ terms and simplifying. The sixth equality re-establishes absolute values, and follows by the defined ordering $A \prec q \prec B \prec C \prec D \prec q' \prec E$. The inequality in the second-last line must be strict since by definition $\Delta > 0$. Finally, the last equality undoes the earlier expansion and collects all eight terms which are, by definition, equal to $\sum_{(r,r')\in\Lambda_0(j)} |x_{rj} - x_{r'j}|$. □

With this technical lemma in hand, we are now ready to show that the restriction that all impression allocations within the same audience segment are the same, i.e., $x_{rj} = x_{ij}\ \forall r \in R_i, j \in J$, can be applied to (P2) without loss of optimality.

PROPOSITION EC.4. *All optimal solutions $(\mathbf{x}^*, \mathbf{y}^*)$ to (P2) satisfy the condition $x^*_{rj} = x^*_{r'j}\ \forall (r, r') \in R_i^2, \forall (i, j) \in \Gamma$, i.e., all impressions within the same audience segment have the same allocation.*

Proof. Assume there exists an optimal solution $(\mathbf{x}, \mathbf{y})$ to (P2) that does not satisfy the stated condition. We will construct an alternative solution $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ which (i) satisfies the stated condition, (ii) is feasible in (P2), and (iii) has a strictly lower (better) objective value, thereby contradicting our original assumption that $(\mathbf{x}, \mathbf{y})$ was optimal. We construct the alternative solution $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ as follows, which involves one or more invocations of the following improvement step.

(Improvement Step) Select an audience segment i and a campaign j for which there exists a pair of impressions (r^1, r^2) within audience segment i that have unequal allocations, i.e., $x_{r^1j} < x_{r^2j}$. Let $\mu_{ij} = (1/|R_i|) \sum_{r\in R_i} x_{rj}$ be the average allocation to campaign j from impressions of audience segment i. Since not all impressions $r \in R_i$ have the same campaign-j allocation x_{rj}, there must exist at least one impression q with a below-average allocation, and another impression q' with an above-average allocation; i.e., $x_{qj} < \mu_{ij} < x_{q'j}$. Let $\Delta = \min\{x_{q'j} - \mu_{ij},\ \mu_{ij} - x_{qj}\}$, and modify the values of the allocation $\mathbf{x}$ such that $x_{q'j} \leftarrow x_{q'j} - \Delta$, and $x_{qj} \leftarrow x_{qj} + \Delta$.

(Termination Criterion) Repeatedly invoke the Improvement Step until $x_{rj} = x_{r'j}\ \forall (r, r') \in R_i^2, \forall (i, j) \in \Gamma$, i.e., all impressions within the same audience segment have the same allocation. Upon termination, denote the resulting allocation as $\hat{\mathbf{x}}$. By construction, this allocation satisfies $\hat{x}_{rj} = \mu_{ij}\ \forall r \in R_i,\ \forall j \in J$.

(Finite Convergence) Termination in a finite number of steps is assured, since (i) at each Improvement Step, at least one impression has its allocation x_{rj} updated to a corresponding mean μ_{ij}, (ii) there are a finite number of impressions and campaigns, and (iii) the means μ_{ij} do not change throughout this procedure.

(Feasibility) We now show that the solution $(\hat{\mathbf{x}}, \mathbf{y})$ is feasible in (P2). First, $\hat{x}_{rj} \geq 0\ \forall (r, j) \in \Lambda$ is assured, since the original solution $(\mathbf{x}, \mathbf{y})$ satisfied $x_{rj} \geq 0\ \forall (r, j) \in \Lambda$, and whenever x_{rj} was decreased, its new level was never set to below μ_{ij}, a non-negative quantity. Next, the left-hand side of the demand constraint has the same value under both solutions $(\mathbf{x}, \mathbf{y})$ and $(\hat{\mathbf{x}}, \mathbf{y})$, since, for each campaign $j \in J$,

$$\sum_{r\in\Lambda(j)} \hat{x}_{rj} = \sum_{i\in\Gamma(j)} \left(\sum_{r\in R_i} \hat{x}_{rj}\right) = \sum_{i\in\Gamma(j)} \left(\sum_{r\in R_i} x_{rj}\right) = \sum_{r\in\Lambda(j)} x_{rj}.$$

The middle equality holds because $\sum_{r\in R_i} x_{rj}$ remains unchanged after each Improvement Step (although two x_{rj} values are updated, the net change to the sum is $\Delta - \Delta = 0$). Since the demand constraint was satisfied by $(\mathbf{x}, \mathbf{y})$, it is also satisfied by $(\hat{\mathbf{x}}, \mathbf{y})$. Finally, since the supply constraint is satisfied by the solution $(\mathbf{x}, \mathbf{y})$, the following simplifications show it must also be satisfied by $(\hat{\mathbf{x}}, \mathbf{y})$:

$$\begin{aligned}&\sum_{j\in\Lambda(r)} x_{rj} \le 1\,\forall r \in R \implies \sum_{j\in\Lambda(r)} x_{rj} \le 1\,\forall r \in R_i, \forall i \in I \implies \sum_{r\in R_i}\sum_{j\in\Lambda(r)} x_{rj} \le |R_i|\,\forall i \in I\\ &\implies \sum_{j\in\Gamma(i)} (1/|R_i|) \sum_{r\in R_i} x_{rj} \le 1\,\forall i \in I \implies \sum_{j\in\Gamma(i)} \mu_{ij} \le 1\,\forall i \in I \implies \sum_{j\in\Gamma(i)} \hat{x}_{ij} \le 1\,\forall i \in I.\end{aligned}$$

(Improved Solution) Lemma EC.2 shows that each Improvement Step strictly lowers (improves) the value of the Gini term in the objective of (P2). Since we perform at least one Improvement Step, the solution $(\hat{\mathbf{x}}, \mathbf{y})$ has a strictly lower (better) value for its Gini term than $(\mathbf{x}, \mathbf{y})$. Moreover, both these solutions share the same $\mathbf{y}$, and thus have the same shortfall cost. Therefore, $(\mathbf{x}, \mathbf{y})$ is clearly suboptimal, which contradicts our initial assumption. This completes our proof. $\square$

C.2. Derivation of Decomposition Method

Having just established that (P2) is the impression-level analog to the audience-segment-based (P-ORIG), we now derive our decomposition method by relaxing (P2) and making a number of observations about the resulting relaxation.

Using β_r as the dual value for impression r's supply constraint, after dualizing the supply constraints, (P2) becomes:

$$\begin{aligned}\text{(P3)}\qquad z^{LB} = \min &\sum_{j\in J,(q,r)\in\Lambda_0(j)} \psi_j |x_{qj} - x_{rj}| + \sum_{j\in J} p_j y_j + \sum_{r\in R} \beta_r \Big(\sum_{j\in\Lambda(r)} x_{rj} - 1\Big)\\ \text{s.t.} &\sum_{r\in\Lambda(j)} x_{rj} + y_j = d_j \qquad \forall j \in J \qquad\qquad \text{(demand)}\\ &x_{rj} \ge 0\,\forall (r,j) \in \Lambda; \qquad y_j \ge 0\,\forall j \in J\end{aligned}$$

Since (P3) is a relaxation of (P2), its optimal value is a lower bound for the optimal value of (P-ORIG), i.e., $z^{LB} \le z^*$. Moreover, (P3) decomposes by campaign and its optimal value can be written as $z^{LB} = -\sum_{r\in R} \beta_r + z_j^{LB}$, where each campaign-$j$ subproblem has optimal value z_j^{LB} and takes the following form:

$$\begin{aligned}\text{(P3-j)}\qquad z_j^{LB} = \min\ &\psi_j \sum_{(q,r)\in\Lambda_0(j)} |x_{qj} - x_{rj}| + p_j y_j + \sum_{r\in\Lambda(j)} \beta_r x_{rj}\\ \text{s.t.} &\sum_{r\in\Lambda(j)} x_{rj} + y_j = d_j \qquad\qquad \text{(demand)}\\ &x_{rj} \ge 0\,\forall r \in \Lambda(j); \qquad y_j \ge 0\end{aligned}$$

Although it is somewhat difficult to characterize the optimal solution to (P3-j) in general, the following proposition characterizes the optimal solution to (P3-j) for the special case where all dual values β_r are zero.

PROPOSITION EC.5. *If $\beta_r = 0$ for all $r \in \Lambda(j)$, then the optimal solution to (P3-j) is $x_{rj} = \theta_j$ for all $r \in \Lambda(j)$, and $y_j = 0$. That is, when all supply constraints of (P2) are non-binding (i.e., $\beta_r = 0$), the optimal solution to (P3-j) is to spread the campaign's impression allocation proportionally across its targeted supply of impressions, and to have no shortfall.*

Proof. The proposed solution is feasible, since the demand constraint holds, i.e., $\sum_{r\in\Lambda(j)} \theta_j + 0 = |\Lambda(j)|\theta_j = \hat{s}_j(d_j/\hat{s}_j) = d_j$. Moreover, this solution has objective value 0, which must be optimal since solutions of negative value are not possible. $\square$

The remainder of this section derives a number of transformations of (P3-j), the last of which allows us to fully characterize the optimal solution of (P3-j). The following theorem and corollary establish a partial characterization for the optimal solution of (P3-j), and are an important step towards fully characterizing the optimal solution of (P3-j).

THEOREM EC.1. *Without loss of generality, order the impressions of $\Lambda(j)$ by β_r such that $\beta_1 \le \beta_2 \le \cdots \le \beta_{\hat{s}_j}$, where $\hat{s}_j = |\Lambda(j)|$. If the solution (x_j^*, y_j^*) where $x_j^* \equiv \{x_{rj}^*\}_{r=1..\hat{s}_j}$ is optimal for (P3-j), then we must have $x_{1j}^* \ge x_{2j}^* \ge \cdots \ge x_{\hat{s}_j j}^*$.*

Proof. Assume for a contradiction that (x_j^*, y_j^*) with $x_j^* \equiv \{x_{rj}^*\}_{r=1..\hat{s}_j}$ is optimal for (P3-j) yet for some pair of impressions (q, q') we have $\beta_q \le \beta_{q'}$ and $x_{qj}^* < x_{q'j}^*$. Let $\Delta = \frac{1}{2}(x_{q'j}^* - x_{qj}^*)$ and construct the solution $(\overline{x}_j, y_j^*)$, where $\overline{x}_j \equiv \{\overline{x}_{rj}\}_{r=1..\hat{s}_j}$, by taking $\overline{x}_{q'j} = x_{q'j}^* - \Delta$, $\overline{x}_{qj} = x_{qj}^* + \Delta$, and $\overline{x}_{rj} = x_{rj}^*\ \forall r \notin \{q, q'\}$. Note that by construction, $\Delta > 0$ and $\overline{x}_{q'j} = \overline{x}_{qj}$. We will now show that $(\overline{x}_j, y_j^*)$ has a lower value than (x_j^*, y_j^*) when evaluated in the objective of (P3-j), contradicting our assumption that (x_j^*, y_j^*) was optimal. First, it is clear that $\sum_{r\in\Lambda(j)} \beta_r \overline{x}_{rj} \le \sum_{r\in\Lambda(j)} \beta_r x_{rj}^*$, since:

$$\begin{aligned}
\sum_{r\in\Lambda(j)} \beta_r \overline{x}_{rj} &= \sum_{r\in\Lambda(j): r\notin\{q,q'\}} \beta_r \overline{x}_{rj} + \beta_q \overline{x}_{qj} + \beta_{q'} \overline{x}_{q'j} \\
&= \sum_{r\in\Lambda(j): r\notin\{q,q'\}} \beta_r x_{rj}^* + \beta_q (x_{qj}^* + \Delta) + \beta_{q'} (x_{q'j}^* - \Delta) \\
&= \sum_{r\in\Lambda(j)} \beta_r x_{rj}^* + \Delta(\beta_q - \beta_{q'}) \\
&\le \sum_{r\in\Lambda(j)} \beta_r x_{rj}^*.
\end{aligned}$$

Moreover, $\sum_{(r,r')\in\Lambda_0(j)} |\overline{x}_{rj} - \overline{x}_{r'j}| < \sum_{(r,r')\in\Lambda_0(j)} |x_{rj}^* - x_{r'j}^*|$, by Lemma EC.2. Thus, $\psi_j \sum_{(r,r')\in\Lambda_0(j)} |\overline{x}_{rj} - \overline{x}_{r'j}| + \sum_j p_j y_j^* + \sum_{r\in\Lambda(j)} \beta_r \overline{x}_{rj} < \psi_j \sum_{(r,r')\in\Lambda_0(j)} |x_{rj}^* - x_{r'j}^*| + \sum_j p_j y_j^* + \sum_{r\in\Lambda(j)} \beta_r x_{rj}^*$, contradicting our assumption that (x_j^*, y_j^*) was optimal. It follows that if the impressions are ordered as $\beta_1 \le \beta_2 \le \cdots \le \beta_{\hat{s}_j}$, then any optimal solution (x_j^*, y_j^*) to (P3-j) must satisfy $x_{1j}^* \ge x_{2j}^* \ge \cdots \ge x_{\hat{s}_j j}^*$. $\square$

COROLLARY EC.2. *Without loss of generality, order the impressions of $\Lambda(j)$ by β_r such that $\beta_1 \leq \beta_2 \leq \cdots \leq \beta_{\hat{s}_j}$, where $\hat{s}_j = |\Lambda(j)|$. Given any two impressions (q, q') with equal dual values $\beta_q = \beta_{q'}$, the optimal allocations for these impressions in (P3-j) are also equal, i.e., $x^*_{qj} = x^*_{q'j}$.*

Proof. Without loss of generality, assume $q' < q$ in the impression ordering. By Theorem EC.1, $x^*_{q'j} \geq x^*_{qj}$. Assume for a contradiction that $x^*_{q'j} > x^*_{qj}$, let $\Delta = \frac{1}{2}(x^*_{q'j} - x^*_{qj})$, and construct the solution $(\overline{x}_j, y^*_j)$, where $\overline{x}_j \equiv \{\overline{x}_{rj}\}_{r=1..m}$, by taking $\overline{x}_{q'j} = x^*_{q'j} - \Delta$, $\overline{x}_{qj} = x^*_{qj} + \Delta$, and $\overline{x}_{rj} = x^*_{rj}\ \forall r \notin \{q, q'\}$. First, it is clear that $\sum_{r\in\Lambda(j)} \beta_r \overline{x}_{rj} = \sum_{r\in\Lambda(j)} \beta_r x^*_{rj}$, since:

$$\begin{aligned}
\sum_{r\in\Lambda(j)} \beta_r \overline{x}_{rj} &= \sum_{r\in\Lambda(j):r\notin\{q,q'\}} \beta_r \overline{x}_{rj} + \beta_q \overline{x}_{qj} + \beta_{q'} \overline{x}_{q'j} \\
&= \sum_{r\in\Lambda(j):r\notin\{q,q'\}} \beta_r x^*_{rj} + \beta_q (x^*_{qj} + \Delta) + \beta_{q'} (x^*_{q'j} - \Delta) \\
&= \sum_{r\in\Lambda(j)} \beta_r x^*_{rj} + \Delta(\beta_q - \beta_{q'}) \\
&= \sum_{r\in\Lambda(j)} \beta_r x^*_{rj},
\end{aligned}$$

where the last line follows since $\beta_q = \beta_{q'}$. Moreover, Lemma EC.2 provides $\sum_{(r,r')\in\Lambda_0(j)} |\overline{x}_{rj} - \overline{x}_{r'j}| < \sum_{(r,r')\in\Lambda_0(j)} |x^*_{rj} - x^*_{r'j}|$. Thus, $\psi_j \sum_{(r,r')\in\Lambda_0(j)} |\overline{x}_{rj} - \overline{x}_{r'j}| + \sum_j p_j y^*_j + \sum_{r\in\Lambda(j)} \beta_r \overline{x}_{rj} < \psi_j \sum_{(r,r')\in\Lambda_0(j)} |x^*_{rj} - x^*_{r'j}| + \sum_j p_j y^*_j + \sum_{r\in\Lambda(j)} \beta_r x^*_{rj}$, implying (x^*_j, y^*_j) is not optimal. This is a contradiction; hence, we conclude that $x^*_{q'j} = x^*_{qj}$. □

Corollary EC.2 is important for two reasons. First, if we restrict ourselves to considering β_r values that are homogenous across all impressions in each viewer type, i.e., $\beta_r = \hat{\beta}_i\ \forall r \in R_i$, then any optimal solution $\{x^*_{rj}\}_{r\in\Lambda(j)}$ to (P3-j) satisfies $x^*_{rj} = x^*_{r'j}\ \forall (r, r') \in R_i^2$ and can be represented in the viewer type space using variables $\{x^*_{ij}\}_{i\in\Gamma(j)}$ such that $x^*_{rj} = x^*_{ij}\ \forall r \in R_i$. Second, we can further exploit this property by aggregating viewer types by dual value $\hat{\beta}_i$, since any viewer types that have the same dual values ($\hat{\beta}_i = \hat{\beta}_{i'}$) must also have equal solutions $x^*_{ij} = x^*_{i'j}$. Finally, we note that we can assume without loss of optimality that β_r values are homogenous across all impressions in each viewer type, since (i) Proposition EC.4 shows that without loss of optimality we can restrict consideration to solutions that satisfy $x_{rj} = x_{r'j}\ \forall (r, r') \in R_i^2$, (ii) under this restriction the objective term $\sum_{r\in R} \beta_r \left(\sum_{j\in\Lambda(r)} x_{rj} - 1\right)$ of (P3-j) simplifies to $\sum_{i\in I} \left(\sum_{r\in R_i} \beta_r\right) \left(\sum_{j\in\Gamma(i)} x_{ij} - 1\right)$, and (iii) in this form, it is clear that the individual values of β_r do not matter, just the sum $\sum_{r\in R_i} \beta_r$, and thus without loss of generality we can assume $\beta_r = \hat{\beta}_i\ \forall r \in R_i$.

Theorem EC.1 is important, as it allows us to simplify (P3-j) in the manner defined by the following proposition.

PROPOSITION EC.6. *Defining $c_{rj} = \psi_j(\hat{s}_j - 2r + 1) + \beta_r$ for all impressions $r \in \Lambda(j)$ ordered according to $\beta_1 \leq \beta_2 \leq \cdots \leq \beta_{\hat{s}_j}$, where $\hat{s}_j = |\Lambda(j)|$, we can write subproblem (P3-j) as follows:*

$$\textit{(P4-j)} \qquad z_j^{LB} = \min \sum_{r=1..\hat{s}_j} c_{rj} x_{rj} + p_j y_j$$

$$\textit{s.t.} \sum_{r=1..\hat{s}_j} x_{rj} + y_j = d_j \qquad \textit{(demand)}$$
$$x_{1j} \geq x_{2j} \geq \cdots \geq x_{\hat{s}_j j} \geq 0; \qquad y_j \geq 0$$

Proof. Theorem EC.1 allows us to impose the optimality cut $x_{1j} \geq x_{2j} \geq \cdots \geq x_{\hat{s}_j j} \geq 0$. These constraints on the variables then allow us to simplify the objective of (P3-j) by removing the absolute values as follows:

$$\psi_j \sum_{(q,r)\in\Lambda_0(j)} |x_{qj} - x_{rj}| + p_j y_j + \sum_{r\in\Lambda(j)} \beta_r x_{rj} = \psi_j \sum_{\substack{q=1..\hat{s}_j-1 \\ r=q+1..\hat{s}_j}} (x_{qj} - x_{rj}) + p_j y_j + \sum_{r=1..\hat{s}_j} \beta_r x_{rj}$$

$$= \psi_j \left(\sum_{\substack{q=1..\hat{s}_j-1 \\ r=q+1..\hat{s}_j}} x_{qj} - \sum_{\substack{r=2..\hat{s}_j \\ q=1..r-1}} x_{rj} \right) + p_j y_j + \sum_{r=1..\hat{s}_j} \beta_r x_{rj}$$

$$= \psi_j \left(\sum_{q=1..\hat{s}_j-1} (\hat{s}_j - q) x_{qj} - \sum_{r=2..\hat{s}_j} (r-1) x_{rj} \right) + p_j y_j + \sum_{r=1..\hat{s}_j} \beta_r x_{rj}$$

$$= \psi_j \left(\hat{s}_j \sum_{q=1..\hat{s}_j-1} x_{qj} - \sum_{q=1..\hat{s}_j-1} q x_{qj} - \sum_{r=2..\hat{s}_j} r x_{rj} + \sum_{r=2..\hat{s}_j} x_{rj} \right) + p_j y_j + \sum_{r=1..\hat{s}_j} \beta_r x_{rj}$$

$$= \psi_j \left(\hat{s}_j \sum_{r=1..\hat{s}_j-1} x_{rj} - \sum_{r=1..\hat{s}_j-1} r x_{rj} - \sum_{r=2..\hat{s}_j} r x_{rj} + \sum_{r=2..\hat{s}_j} x_{rj} \right) + p_j y_j + \sum_{r=1..\hat{s}_j} \beta_r x_{rj}$$

$$= \psi_j \left(\hat{s}_j \left(x_{1j} + \sum_{r=2..\hat{s}_j-1} x_{rj} \right) - \left(x_{1j} + \sum_{r=2..\hat{s}_j-1} r x_{rj} \right) \right.$$
$$\left. - \left(\sum_{r=2..\hat{s}_j-1} r x_{rj} + \hat{s}_j x_{\hat{s}_j j} \right) + \left(\sum_{r=2..\hat{s}_j-1} x_{rj} + x_{\hat{s}_j j} \right) \right) + p_j y_j + \sum_{r=1..\hat{s}_j} \beta_r x_{rj}$$

$$= \psi_j \left((\hat{s}_j - 1) x_{1j} + \sum_{r=2..\hat{s}_j-1} (\hat{s}_j - 2r + 1) x_{rj} + (-\hat{s}_j + 1) x_{\hat{s}_j j} \right) + p_j y_j + \sum_{r=1..\hat{s}_j} \beta_r x_{rj}$$

$$= \psi_j \left(\sum_{r=1..\hat{s}_j} (\hat{s}_j - 2r + 1) x_{rj} \right) + p_j y_j + \sum_{r=1..\hat{s}_j} \beta_r x_{rj} = \sum_{r=1..\hat{s}_j} c_{rj} x_{rj} + p_j y_j,$$

as required. □

Next, the following proposition describes how to aggregate (P4-j) to the viewer type space. Further simplifications of (P4-j) will be presented in the viewer type space, allowing us to read off our final result in the viewer type space.

PROPOSITION EC.7. *Without loss of generality, order all viewer types $i \in \Gamma(j)$ according to $\hat{\beta}_1 \leq \hat{\beta}_2 \leq \cdots \leq \hat{\beta}_{m_j}$, where $m_j = |\Gamma(j)|$. Define $c_{ij} = \psi_j s_i (s_i^{aj} - s_i^{bj}) + s_i \hat{\beta}_i$, where $s_i^{bj} = \sum_{i'=1..i-1} s_{i'}$ and $s_i^{aj} = \sum_{i'=i+1..m_j} s_{i'}$ are the number of impressions that are rank ordered before and after viewer type i's impressions, respectively (thus, $\hat{s}_j = s_i^{bj} + s_i + s_i^{aj}$). Taking $\beta_r = \hat{\beta}_i \; \forall r \in R_i$, we represent subproblem (P4-j) in the viewer type space by aggregating all impressions of each viewer type together. The resulting formulation is:*

$$\begin{aligned}
\textit{(P5-j)} \qquad z_j^{LB} = \min & \sum_{i=1..m_j} c_{ij} x_{ij} + p_j y_j \\
\textit{s.t.} & \sum_{i=1..m_j} s_i x_{ij} + y_j = d_j && \textit{(demand)} \\
& x_{1j} \geq x_{2j} \geq \cdots \geq x_{m_j j} \geq 0; \qquad y_j \geq 0
\end{aligned}$$

Proof. By Corollary EC.2, since $\beta_r = \hat{\beta}_i \; \forall r \in R_i$, we must have $x_{rj} = x_{ij} \; \forall r \in R_i$. The left-hand-side of the demand constraint aggregates as follows:

$$\sum_{r \in \Lambda(j)} x_{rj} = \sum_{i \in \Gamma(j)} \sum_{r \in R_i} x_{rj} = \sum_{i \in \Gamma(j)} \sum_{r \in R_i} x_{ij} = \sum_{i \in \Gamma(j)} |R_i| x_{ij} = \sum_{i=1..m_j} s_i x_{ij}.$$

Moreover, for each viewer type i we have one term in the objective which was aggregated from the impression space as follows. Let us assume that viewer type i corresponds to the impressions $\{q, \ldots, q'\}$. Therefore, by construction, $s_i = q' - q + 1$ is the size of viewer type i created by aggregating impressions $\{q, \ldots, q'\}$ together. By definition, we have:

$$\begin{aligned}
\sum_{r=q..q'} c_{rj} x_{rj} &= \sum_{r=q..q'} \left(\psi_j (\hat{s}_j - 2r + 1) + \beta_r\right) x_{rj} \\
&= \left(\psi_j \sum_{r=q..q'} (\hat{s}_j - 2r + 1) + \sum_{r=q..q'} \beta_r \right) x_{ij} \\
&= \left(\psi_j \left(\sum_{r=q..q'} (\hat{s}_j + 1) - 2 \sum_{r=q..q'} r \right) + \sum_{r=q..q'} \hat{\beta}_i \right) x_{ij} \\
&= \left(\psi_j \left(s_i(\hat{s}_j + 1) - 2\left(s_i(q + q')/2\right)\right) + s_i \hat{\beta}_i\right) x_{ij} \\
&= \left(\psi_j s_i \left(\hat{s}_j + 1 - q - q'\right) + s_i \hat{\beta}_i\right) x_{ij} \\
&= \left(\psi_j s_i \left((\hat{s}_j - q') - (q - 1)\right) + s_i \hat{\beta}_i\right) x_{ij} \\
&= \left(\psi_j s_i \left(s_i^{aj} - s_i^{bj}\right) + s_i \hat{\beta}_i\right) x_{ij} = c_{ij} x_{ij},
\end{aligned}$$

as required. □

Next, we reformulate (P5-j) using the following variable transformation, which will make it easier to read off the optimal solution.

PROPOSITION EC.8. *Given viewer types $i \in \Gamma(j)$ ordered according to $\hat{\beta}_1 \leq \hat{\beta}_2 \leq \cdots \leq \hat{\beta}_{m_j}$, where $m_j = |\Gamma(j)|$, define $\tilde{c}_{ij} = \sum_{i'=1..i} c_{i'j}$ and $\tilde{s}_{ij} = \sum_{i'=1..i} s_{i'}$. Let $\delta_{m_j j} = x_{m_j j}$, and $\delta_{ij} = x_{ij} - x_{i+1,j}$ for all $i = 1..m_j - 1$. We can reformulate (P5-j) using the decision variables $\{\delta_{ij}\}_{i=1..m_j}$ as follows:*

$$\text{(P6-j)} \qquad z_j^{LB} = \min \sum_{i=1..m_j} \tilde{c}_{ij}\delta_{ij} + p_j y_j$$

$$\text{s.t.} \sum_{i=1..m_j} \tilde{s}_{ij}\delta_{ij} + y_j = d_j \qquad \text{(demand)}$$

$$\delta_{ij} \geq 0 \; \forall i = 1..m_j; \; y_j \geq 0$$

Proof. By definition, $x_{ij} = \sum_{i'=i..m_j} \delta_{i'j}$. The left-hand side of the demand constraint from (P5-j) transforms via substitution as follows:

$$\sum_{i=1..m_j} s_i x_{ij} = \sum_{i=1..m_j} s_i \left(\sum_{i'=i..m_j} \delta_{i'j} \right) = \sum_{i'=1..m_j} \delta_{i'j} \left(\sum_{i=1..i'} s_i \right) = \sum_{i'=1..m_j} \tilde{s}_{i'j}\delta_{i'j}.$$

Moreover, the constraints $x_{ij} \geq x_{i+1,j}$ for $i = 1..m_j - 1$ in (P5-j) imply $x_{ij} - x_{i+1,j} \geq 0$, i.e., $\delta_{ij} \geq 0$ for $i = 1..m_j - 1$. And the constraint $x_{m_j j} \geq 0$ implies $\delta_{m_j j} \geq 0$. Finally, the objective of (P5-j) transforms via substitution as follows:

$$\sum_{i=1..m_j} c_{ij} x_{ij} = \sum_{i=1..m_j} c_{ij} \left(\sum_{i'=i..m_j} \delta_{i'j} \right) = \sum_{i'=1..m_j} \delta_{i'j} \left(\sum_{i=1..i'} c_{ij} \right) = \sum_{i'=1..m_j} \tilde{c}_{i'j}\delta_{i'j}. \qquad \square$$

We can now read off the optimal solution of (P6-j) in the "increment space" defined by the δ_{ij} variables.

PROPOSITION EC.9. *Assume viewer types $i \in \Gamma(j)$ are ordered according to the supply duals $\hat{\beta}_1 \leq \hat{\beta}_2 \leq \cdots \leq \hat{\beta}_{m_j}$, where $m_j = |\Gamma(j)|$. Let $\pi_{ij} = \tilde{c}_{ij}/\tilde{s}_{ij} \; \forall i = 1..m_j$ and define $i^* = \arg\min_{i \in \{1..m_j\}} \pi_{ij}$. Then, the optimal solution and value to (P6-j) take the following form. If $\pi_{i^*j} > p_j$, then $y_j^* = d_j$, $\delta_{ij}^* = 0 \; \forall i = 1..m_j$, with corresponding optimal value $p_j d_j$. Otherwise, the optimal value is $\pi_{i^*j} d_j$ with corresponding optimal solution $y_j^* = 0$, and*

$$\delta_{ij}^* = \begin{cases} d_j/\tilde{s}_{ij} & \text{for } i = i^* \\ 0 & \text{for } i \neq i^* \end{cases}$$

Proof. The δ_{ij} variable with the lowest cost per unit of demand satisfied is indexed by i^*. If this cheapest variable is too expensive ($\pi_{i^*j} > p_j$), we prefer not to satisfy this campaign at all, and set shortfall equal to demand, i.e., $y_j^* = d_j$. On the other hand, if using this cheapest variable is no more expensive than incurring shortfall ($\pi_{i^*j} \leq p_j$), we satisfy demand fully with this variable by using the solution $\{\delta_{ij}^*\}_{i=1..m_j}$ as defined. In the former case, the optimal value is $p_j y_j^* = p_j d_j$. In the latter case, the optimal value is $\tilde{c}_{i^*j}\delta_{i^*j}^* = \tilde{c}_{i^*j}(d_j/\tilde{s}_{i^*j}) = (\tilde{c}_{i^*j}/\tilde{s}_{i^*j})d_j = \pi_{i^*j} d_j$. $\square$

Finally, we can translate the solution represented by the δ-variables in the so-called "increment space" back to the original space represented by x-variables. The following theorem summarizes our main structural result.

THEOREM EC.2. *Assume viewer types $i \in \Gamma(j)$ are ordered according to the supply duals $\hat{\beta}_1 \le \hat{\beta}_2 \le \cdots \le \hat{\beta}_{m_j}$, where $m_j = |\Gamma(j)|$. Let $\pi_{ij} = \tilde{c}_{ij}/\tilde{s}_{ij}\ \forall i = 1..m_j$ and define $i^* = \arg\min_{i\in\{1..m_j\}} \pi_{ij}$. Then, the optimal solution to (P5-j) takes the following form. If $\pi_{i^*j} > p_j$, then $y_j^* = d_j$ and $x_{ij}^* = 0\ \forall i = 1..m_j$. Otherwise, $y_j^* = 0$ and*

$$x_{ij}^* = \begin{cases} d_j/\tilde{s}_{i^*j} & \textit{for } i \le i^* \\ 0 & \textit{for } i > i^* \end{cases}$$

Proof. From Proposition EC.8, we know that $x_{ij}^* = \sum_{i'=i..m_j} \delta_{i'j}^*$. Combining this fact with the optimal solution $\{\delta_{ij}^*\}_{i=1..m_j}$ from Proposition EC.9 yields the desired result. □

Theorem EC.2 fully characterizes the solution to (P5-j), and equivalently to (P3-j). If shortfall costs are relatively low ($p_j < \pi_{i^*j}$), then it is optimal to allocate no impressions to campaign j, and incur full shortfall ($y_j^* = d_j$). If shortfall costs are relatively high ($p_j \ge \pi_{i^*j}$), then it is optimal to have no shortfall ($y_j^* = 0$) and to spread impressions proportionally across the viewer types $\{1,\ldots,i^*\}$. Recall that since viewer types $i \in \Gamma(j)$ are ordered according to the supply duals $\hat{\beta}_1 \le \hat{\beta}_2 \le \cdots \le \hat{\beta}_{m_j}$, where $m_j = |\Gamma(j)|$, and so it is possible to interpret the solution as one that proportionally spreads impressions across the cheapest set of viewer types $\{1,\ldots,i^*\}$. The viewer type i^* is at the threshold of affordability, given the values for $\{\hat{\beta}_i\}_{i=1..m_j}$.

We can now solve (P3-j) efficiently by sorting the viewer types in order of $\hat{\beta}_i$, and then invoking Theorem EC.2 to compute the optimal solution to (P3-j) analytically. Notice that Theorem 1 of §6 is essentially Theorem EC.2 with a number of parameters which we formally defined in previous propositions now written out in the statement of the theorem. The full decomposition method, which involves a way to choose $\hat{\beta}_i$'s as well as a way to produce a near-feasible solution to (P-ORIG) from one or more optimal solutions of (P3-j) computed from different $\{\hat{\beta}_i\}_{i\in I}$ values, is described in §6.

Appendix D: Online Algorithm

While the solution to our Gini-based model (SG) can be used directly to serve ads in real-time by periodically re-solving (SG) over a rolling horizon as supply forecasts are updated, and for each arriving user of type i serving ad $j \in \Gamma(i)$ with probability x_{ij}, we can also use information generated by our decomposition method to construct an online algorithm that adapts the ad to show based on the types of impressions which actually materialize. This approach is in line with that of Devanur and Hayes (2009), Feldman et al. (2010), Vee et al. (2010), Mehta (2012), and Agrawal et al. (2014), who have developed online algorithms for a variety of different (non-Gini-based) ad allocation problems. In what follows, we provide a sketch of an online algorithm whose structure is in line with our Gini-based model and associated decomposition method. Since a full formal analysis of the behavior of an online algorithm can be quite lengthy, we leave this to future work. Nevertheless, the following description is useful to draw structural parallels between our decomposition method and the online algorithm it suggests.

Recall that our decomposition method (Section 6) produces the following outputs:

- A vector $\lambda_j = \{\lambda_{j0}, \lambda_{j1}, \ldots, \lambda_{jN}\}$ for each campaign j whose elements λ_{jn}, $n = 0..N$, are all non-negative and satisfy $\sum_{n=0..N} \lambda_{jn} = 1$. That is, λ_j is a probability vector.
- A vector $\hat{\beta}^n = \{\hat{\beta}_1^n, \hat{\beta}_2^n, \ldots, \hat{\beta}_m^n\}$ for each iteration $n = 0..N$ whose elements $\hat{\beta}_i^n$, $i \in I \equiv \{1..m\}$, are the dual prices computed in iteration n for each audience segment i. That is, $\hat{\beta}^n$ is a price vector.

We propose the following online algorithm that decides, on-the-fly, what ad to show to each arriving user.

(Initialization) Set $x_{ij}^S := 0$ for all $(i,j) \in \Gamma$. We will use x_{ij}^S to keep track of any surplus allocation from one arrival to the next (the 'S' superscript denotes "surplus").

(Repeat) For each arriving user,

1. Identify the user as a member of a particular audience segment i, and loop through all campaigns $j \in \Gamma(i)$ which target this audience segment, performing the following:

   (a) Randomly draw $n' := n$ with probability λ_{jn}, and let $\hat{\beta} := \hat{\beta}^{n'}$ be the price vector that campaign j faces. That is, campaign j faces the price vector $\hat{\beta}^n$ with probability λ_{jn}.

   (b) Invoke Theorem 1 using price vector $\hat{\beta}$ to compute x_{ij}.

2. At this stage, we have a proposed allocation x_{ij}, $j \in \Gamma(i)$. Augment this allocation with the surplus allocation from previous arrivals of the same type, i.e., compute $x_{ij}^A := x_{ij} + x_{ij}^S$ for all $j \in \Gamma(i)$.

3. Compute $\xi := 1/\max(1, \sum_{j\in\Gamma(i)} x_{ij}^A)$. Then for all $j \in \Gamma(i)$, let $x_{ij}^P := \xi x_{ij}^A$ and $x_{ij}^S := (1-\xi)x_{ij}^A$. Notice that, by construction, $\xi \in [0,1]$ and $\sum_{j\in\Gamma(i)} x_{ij}^P \leq 1$.

4. Show ad $j \in \Gamma(i)$ with probability x_{ij}^P, and show no ad with probability $1 - \sum_{j\in\Gamma(i)} x_{ij}^P$. Any surplus allocation x_{ij}^S is carried forward to the next user arrival of type i.

By design, this online algorithm over-serves excess impressions to a campaign if the supply matching a campaign exceeds the forecasted traffic volume. On the other hand, if we wish to avoid over-serving, then we can keep track of the total number of impressions delivered to each campaign and modify step 1b so that Theorem 1 is invoked to compute x_{ij} when fewer than the demanded d_j impressions have been served to campaign j, and $x_{ij} := 0$ once the campaign's demand d_j is satisfied.

The structure of our online algorithm follows directly from the Dantzig-Wolfe master problem (PM). Indeed, we are reconstructing, on-the-fly, all solutions x_{ij}^n, $n = 0..N$, according to the probabilities λ_{jn}, and randomizing over these solutions as the Dantzig-Wolfe master problem dictates. By construction, if our online algorithm is given a sufficiently large number of user arrivals of type i, then by the law of large numbers the actual proportion of arrivals of type i which see the solution x_{ij}^n converges to λ_{jn}, and thus with probability 1 the resulting allocation is feasible (i.e., satisfies the supply constraint in (PM)). Moreover, if we get exactly s_i arrivals of each type i, then by a similar asymptotic argument the value achieved by our online algorithm as measured by our Gini objective (6) converges to the master problem's optimal value. It follows that if our decomposition method is run to optimality, and its inputs are used by our online algorithm, then our online algorithm is asymptotically optimal when the number of arrivals of each type are equal to their forecasted quantities. More generally, if our decomposition method is terminated early with a measurable optimality gap, i.e., with N smaller than it would be if the decomposition was run to completion, and/or our forecasts for the supplies s_i have some error, then our online algorithm produces a robust yet near-optimal solution.

References

Agrawal S, Wang Z, Ye Y (2014) A dynamic near-optimal algorithm for online linear programming. *Operations Research* 62(4): 876–890.

Devanur NR, Hayes TP (2009) The adwords problem: online keyword matching with budgeted bidders under random permutations. In *Proc. 10th ACM Conf. Electronic Commerce, EC'09*, New York, NY, pp. 71–78.

Feldman J, Henzinger M, Korula N, Mirrokni V, Stein C (2010) Online stochastic packing applied to display ad allocation. In *Proc. of European Symposium on Algorithms*, Liverpool, UK, pp. 182–194.

Mehta A (2012) Online matching and ad allocation. *Theoretical Comput. Sci.* 8(4): 265–368.

Vee E, Vassilvitskii S, Shanmugasundaram J (2010) Optimal online assignment with forecasts. Parkes DC, Dellarocas C, Tennenholtz M, eds. In *Proc. 11th ACM Conf. Electronic Commerce, EC'10*, New York, NY, 109–118.

Appendix E: Multi-Period Model

In this section, we extend our single-period Gini-based ad planning model (SG) to a multi-period (and thus, multi-dimensional) model. We begin in §E.1 with a short literature review of models that spread impressions across time, as well as those that spread across both audience segments and time. We then introduce a multi-period baseline model (§E.2), define relevant Gini-based metrics (§E.3), use these metrics to formulate a multi-period Gini model (§E.4), and illustrate the Lorenz curves produced by our multi-period model (§E.5). We then introduce a novel decomposition method for our multi-period Gini problem that nests the single-period decomposition method of §6 into a Lagrangian Decomposition scheme (§E.6), and conclude by demonstrating tail Gini metrics that may be used to soften the penalties for deviating from impression targets (§E.7).

E.1. Multi-Period Spreading Literature

As we have described in §2.1, advertisers prefer for the impressions that they get to be well-spread across audience segments. But, this is not the only dimension on which advertisers care to have their impressions spread. In practice, advertisers also often want their campaigns to be well-paced, i.e., with impressions delivered evenly across time. Pacing ensures ads reach a wide audience, have a sustained impact, and support complementary ads that are being run in other media (e.g., television). Bhalgat et al. (2012) use several nested packing constraints to develop a tight $(1-1/e)$-competitive online algorithm for pacing impressions over time. Meanwhile, Lee et al. (2013) and Xu et al. (2015) employ control-based heuristics to throttle the rate impressions are purchased from a spot market to satisfy guaranteed campaigns, with Lee et al. (2013) adjusting the value of bids over time and Xu et al. (2015) keeping bids constant but instead throttling the auction participation rate (both have been tested with real data, and the latter describes a real implementation). Araman and Fridgeirsdottir (2015) use a queuing model and fluid analysis to determine asymptotically optimal prices and display frequencies for well-paced ads.

Finally, Bollapragada et al. (2002) and Turner et al. (2011) have studied the problem of spreading impressions across *both* audience segments and time, for television and dynamic in-game advertising, respectively. Methodologically, all of these papers are different than ours, since they do not consider Gini metrics. To the best of our knowledge, ours is the first paper to propose using Gini metrics to spread online ad impressions.

E.2. Multi-Period Baseline Model

We use the model of Turner et al. (2011), developed for Microsoft (then Massive, Inc.) to plan and schedule dynamic in-game advertising, as our baseline multi-period online ad planning model. In this model, each advertiser defines, in its contract with the publisher, a set of goals that should be met, along with weights that capture the relative importance of each goal. The primary goal is always the end-of-campaign impression goal, which, as in (SB) and (SG), states that each campaign j should get a total of d_j impressions by its end date, otherwise the shortfall y_j is penalized linearly in the objective. In addition, there are a number of secondary goals that ensure impressions are also well-spread (i) over targeted audience segments, (ii) over time, and (iii) over targeted audience segments at each point in time. Note that because (iii) is the most difficult to achieve, criteria (i) and (ii) are given more weight, and thus prioritized ahead of (iii). Consequently, (i) and (ii) are not subsumed by (iii).

Formally, spreading goals (i), (ii), and (iii) are modeled in the following general way. First, each goal is expressed as an impression target d. Then, lower and upper bounds (ℓ, u) are defined which bracket the impression target d; i.e., $\ell \leq d \leq u$. Oftentimes, the bounds are defined in a symmetric manner, e.g., $\ell = \alpha d$ and $u = (1/\alpha)d$, for $\alpha \in [0,1]$; however, that need not be the case. As long as the number of allocated impressions w falls within the bounds, this is perceived as "good enough spreading", and the publisher is not penalized. However, if w is above or below the bounds, then the extent to which the allocation falls outside of the bounds is penalized linearly. Formally, the objective function minimizes a sum of terms of the form py, where y is the extent to which the bounds are violated and is a decision variable. The impression allocation $w = sx$ is equal to the supply of impressions s multiplied by the proportion of impressions allocated x. Constraints of the form $\ell - y \leq w \leq u + y$; $w \geq 0$; $y \geq 0$ link w with y. We use various forms of x and y as decision variables, and only implicitly model $w \equiv sx$. Indexing time periods with $t \in T$, we get a model with the essential structure of Turner et al. (2011), that captures the primary impression goals (modeled by the demand constraints) as well as the secondary spreading goals of types (i), (ii), and (iii). The full model, which uses variables x and y, as well as parameters p, d, s, ℓ, and u with different subscripts as appropriate from the context, is as follows:

$$\text{(MB)} \qquad \min \sum_{j \in J} p_j y_j + \sum_{(i,j) \in \Gamma} p_{ij} y_{ij} + \sum_{j \in J, t \in T_j} p_{jt} y_{jt} + \sum_{(i,j) \in \Gamma, t \in T_j} p_{ijt} y_{ijt} \tag{EC.5a}$$

$$\text{s.t.} \quad \sum_{i \in \Gamma(j), t \in T_j} s_{it} x_{ijt} + y_j = d_j \qquad \forall j \in J \qquad \text{(demand)} \tag{EC.5b}$$

$$\sum_{j \in \Gamma(i) \cap J_t} x_{ijt} \leq 1 \qquad \forall i \in I, t \in T \qquad \text{(supply)} \tag{EC.5c}$$

$$\ell_{ij} - y_{ij} \leq \sum_{t \in T_j} s_{it} x_{ijt} \leq u_{ij} + y_{ij} \qquad \forall (i,j) \in \Gamma \qquad \text{(spreading type (i))} \tag{EC.5d}$$

$$\ell_{jt} - y_{jt} \leq \sum_{i \in \Gamma(j)} s_{it} x_{ijt} \leq u_{jt} + y_{jt} \qquad \forall j \in J, t \in T_j \qquad \text{(spreading type (ii))} \tag{EC.5e}$$

$$\ell_{ijt} - y_{ijt} \leq s_{it} x_{ijt} \leq u_{ijt} + y_{ijt} \qquad \forall (i,j) \in \Gamma, t \in T_j \qquad \text{(spreading type (iii))} \tag{EC.5f}$$

$$x_{ijt}, y_j, y_{ij}, y_{jt}, y_{ijt} \geq 0 \qquad \forall (i,j) \in \Gamma, t \in T_j \qquad \text{(non-negativity)} \tag{EC.5g}$$

The impression targets, i.e., sub-demands, for the three spread constraints are defined as (i) $d_{ij} = s_i \theta_j = \frac{s_i}{\hat{s}_j} d_j$, (ii) $d_{jt} = (1/|T_j|) d_j$, and (iii) $d_{ijt} = \frac{s_{it}}{\hat{s}_{jt}} d_j$, respectively, where T_j represents the set of time periods which campaign j spans, J_t represents the campaigns active in period t, and $\hat{s}_{jt} = \sum_{i \in \Gamma(j)} s_{it}$ is the number of impressions eligible for campaign j at time t. Thus, reasonable choices for bounds are $\ell_{ij} = 0.9 d_{ij}$, $u_{ij} = 1.1 d_{ij}$, $\ell_{jt} = 0.9 d_{jt}$, $u_{jt} = 1.1 d_{jt}$, $\ell_{ijt} = 0.9 d_{ijt}$, and $u_{ijt} = 1.1 d_{ijt}$, assuming we allow a deviation from all impression targets of ten percent.

E.3. Multi-Period Gini-Based Metrics

We now define Gini-based metrics for the three spread dimensions. In practice, spreading over audience segments is defined proportionally, whereas spreading over time is defined using absolute impressions; therefore, the Gini metric for spread type (ii) takes a different form than (i) and (iii). In what follows, T_j is the set of time periods campaign j spans; J_t is the set of campaigns active in period t; s_{it} is the impression supply of audience segment i in time period t; and $\hat{s}_{jt} = \sum_{i\in\Gamma(j)} s_{it}$ is the number of impressions eligible for campaign j at time t. Our primary decision variables are x_{ijt}, the proportion of audience segment i to assign campaign j in time period t.

The Gini-based metric for dimension (i), which spreads impressions proportionally across audience segments, is the same as in the single-period model (c.f., Eq. (5)). Analogously, the Gini-based metric for dimension (iii), which spreads impressions proportionally across audience segments within a single time period, is:

$$G_{jt} = \frac{GMD_{jt}}{2\mu_{jt}} = \frac{\frac{2}{\hat{s}_{jt}^2} \sum\limits_{(h,i)\in\Gamma_0(j)} s_{ht}s_{it}|x_{hjt} - x_{ijt}|}{2\left(\frac{1}{\hat{s}_{jt}}\sum_{i\in\Gamma(j)} s_{it}x_{ijt}\right)} = \left(\frac{1}{\hat{s}_{jt}}\right) \frac{\sum\limits_{(h,i)\in\Gamma_0(j)} s_{ht}s_{it}|x_{hjt} - x_{ijt}|}{\sum_{i\in\Gamma(j)} s_{it}x_{ijt}}. \tag{EC.6}$$

In contrast, the Gini metric for constraint (ii) that spreads impressions across time follows from (1) and, using $T_0(j) = \{(t,\tau)\in T_j^2 : t < \tau\}$, is defined as:

$$G_j^T = \frac{GMD_j^T}{2\mu_j^T} = \frac{\frac{2}{|T_j|^2} \sum\limits_{(t,\tau)\in T_0(j)} |w_{jt} - w_{j\tau}|}{2\left(\frac{1}{|T_j|}w_j\right)} = \left(\frac{1}{|T_j|w_j}\right) \sum_{(t,\tau)\in T_0(j)} |w_{jt} - w_{j\tau}|, \tag{EC.7}$$

where $w_{jt} = \sum_{i\in\Gamma(j)} s_{it}x_{ijt}$ and $w_j = \sum_{t\in T_j} w_{jt}$ are the number of impressions allocated to campaign j in period t and over the planning horizon, respectively. This Gini metric strives for solutions that have the same absolute impression allocation each time period, e.g., 100,000 impressions each week, regardless of whether impression supply $\{\hat{s}_{jt}, t\in T_j\}$ is uniform across time. In contrast, we could define the alternate Gini metric:

$$G_j^{ALT} = \frac{GMD_j^{ALT}}{2\mu_j^{ALT}} = \frac{\frac{2}{\hat{s}_j^2} \sum\limits_{(t,\tau)\in T_0(j)} \hat{s}_{jt}\hat{s}_{j\tau}|x_{jt} - x_{j\tau}|}{2\left(\frac{1}{\hat{s}_j}w_j\right)} = \frac{1}{\hat{s}_j w_j} \sum_{(t,\tau)\in T_0(j)} \hat{s}_{jt}\hat{s}_{j\tau}|x_{jt} - x_{j\tau}|, \tag{EC.8}$$

where $x_{jt} = w_{jt}/\hat{s}_{jt}$ is the proportion of eligible impressions in time period t allocated to campaign j. If $\hat{s}_{j2}$ is twice that of $\hat{s}_{j1}$, this alternate Gini metric will strive to allocate twice as many impressions to the second period as the first. This alternate metric is akin to the proportional spreading that we have defined for constraints (i) and (iii); however, in practice advertisers care more about the absolute spread over time and so we use (EC.7).

E.4. Multi-Period Gini-Based Model

We now introduce our multi-period Gini-based model. Following the spirit of our single-period Gini model (SG), we propose the following objective function:

$$\min\ \alpha_1 \sum_{j\in J} w_j G_j + \alpha_2 \sum_{j\in J} w_j G_j^T + \alpha_3 \sum_{j\in J, t\in T_j} w_{jt}G_{jt} + \sum_{j\in J} p_j y_j, \tag{EC.9}$$

which, defining the total Gini penalty $G_{jt}^{TOTAL} = \alpha_1 G_j + \alpha_2 G_j^T + \alpha_3 G_{jt}$, may be expressed as:

$$\min \sum_{j\in J, t\in T_j} G_{jt}^{TOTAL} w_{jt} + \sum_{j\in J} p_j y_j.$$

This objective again has the useful interpretation (c.f., Property 3) of assigning a cost to each impression demanded by each campaign j, since $\sum_{t\in T_j} w_{jt} + y_j = d_j$. Unallocated impressions (corresponding to a demand shortfall) incur the cost p_j, while impressions allocated to time period t incur the cost G_{jt}^{TOTAL}, which can be viewed as a charge for less-than-perfect service (spreading). Putting all the pieces together yields the following Gini-based linear program[9] over the variables x_{ij}, x_{ijt}, x_{hij}^{+}, x_{hijt}^{+}, w_{jt}, $w_{jt\tau}^{+}$, and y_j:

$$
\begin{aligned}
\text{(MG)} \qquad \min\ & \alpha_1 \sum_{j\in J} \frac{1}{\hat{s}_j} \sum_{(h,i)\in\Gamma_0(j)} s_h s_i x_{hij}^{+} + \alpha_2 \sum_{j\in J} \frac{1}{|T_j|} \sum_{(t,\tau)\in T_0(j)} w_{jt\tau}^{+} \\
& + \alpha_3 \sum_{j\in J, t\in T_j} \frac{1}{\hat{s}_{jt}} \sum_{(h,i)\in\Gamma_0(j)} s_{ht} s_{it} x_{hijt}^{+} + \sum_{j\in J} p_j y_j && \text{(EC.10a)} \\
\text{s.t. }\ & (8a), (8b) && \text{(linking)} \\
& \sum_{i\in\Gamma(j), t\in T_j} s_{it} x_{ijt} + y_j = d_j \qquad \forall j\in J && \text{(demand)} \quad \text{(EC.10b)} \\
& \sum_{j\in\Gamma(i)\cap J_t} x_{ijt} \le 1 \qquad \forall i\in I, t\in T && \text{(supply)} \quad \text{(EC.10c)} \\
& s_i x_{ij} = \sum_{t\in T_j} s_{it} x_{ijt} \qquad \forall (i,j)\in\Gamma && \text{(linking)} \quad \text{(EC.10d)} \\
& w_{jt} = \sum_{i\in\Gamma(j)} s_{it} x_{ijt} \qquad \forall j\in J, \forall t\in T_j && \text{(linking)} \quad \text{(EC.10e)} \\
& x_{hijt}^{+} \ge x_{hjt} - x_{ijt} \qquad \forall j\in J, \forall t\in T_j, \forall (h,i)\in\Gamma_0(j) && \text{(linking)} \quad \text{(EC.10f)} \\
& x_{hijt}^{+} \ge x_{ijt} - x_{hjt} \qquad \forall j\in J, \forall t\in T_j, \forall (h,i)\in\Gamma_0(j) && \text{(linking)} \quad \text{(EC.10g)} \\
& w_{jt\tau}^{+} \ge w_{jt} - w_{j\tau} \qquad \forall j\in J, \forall (t,\tau)\in T_0(j) && \text{(linking)} \quad \text{(EC.10h)} \\
& w_{jt\tau}^{+} \ge w_{j\tau} - w_{jt} \qquad \forall j\in J, \forall (t,\tau)\in T_0(j) && \text{(linking)} \quad \text{(EC.10i)} \\
& x_{ijt}, y_j \ge 0 \qquad \forall j\in J, \forall t\in T_j, \forall (h,i)\in\Gamma_0(j) && \text{(non-negativity)} \quad \text{(EC.10j)}
\end{aligned}
$$

Constraint (EC.10d) links the proportions of supply allocated in each period with the proportions of supply used across all periods, and makes use of the fact that $s_i = \sum_{t\in T_j} s_{it}$. Constraint (EC.10e) links the auxiliary w_{jt} variables to x_{ijt} variables; we note that it is possible to eliminate this constraint as well as the w_{jt} variables by substituting it into (EC.10h) and (EC.10i).

We also point out that (MB) is somewhat less restrictive than (MG), since for each impression target in (MB) there was always a range around the target over which no penalty was incurred. If one would prefer to mimic a similar logic in a Gini-based model, one can generalize the Gini-based metrics defined here to their tail-Gini equivalents, as we illustrate in Appendix E.7.

E.5. Lorenz Curves of Multi-Period Model

Figure EC.4 plots three campaigns (one per row) as well as six Lorenz curves per campaign (one per column), to illustrate the output from our multi-period model (MG). Each Lorenz curve corresponds to a single Gini metric. The first column corresponds to G_j, the Gini metric which measures spreading of impressions across audience segments for the entire time horizon; columns 2-5 correspond to G_{jt} for the time periods $t = 1..4$, which measure the spreading of impressions across audience segments for each time period $t = 1..4$; and the last

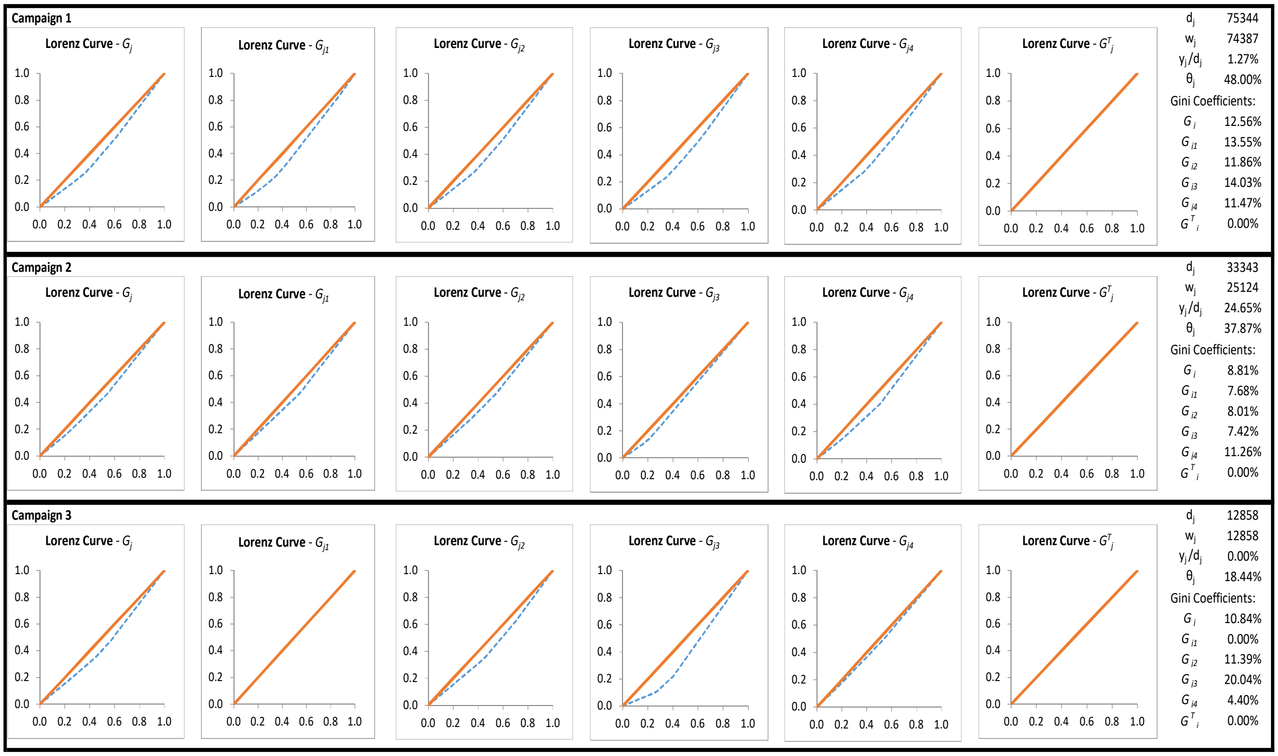

Figure EC.4 Lorenz curves for three campaigns and six Gini metrics.

column corresponds to G_j^T, which measures the spreading of impressions across time. Each advertiser would be able to see only its own Lorenz curves, i.e., a single row of Figure EC.4.

Each panel of our figure illustrates a single Lorenz curve, corresponding to a single Gini metric for a single campaign. The Lorenz curve itself is the dotted blue line, while the orange $45°$ line represents what the Lorenz curve would look like if impressions were perfectly spread. As we can see, in this instance, we are able to achieve perfect spreading across time (see column 6), as well as perfect spreading across audience segments for campaign 3 in week 1. The worst spreading occurs for campaign 3 in week 3, which is reflected by the highest G_{jt} value. The area between the Lorenz curve and its nominal $45°$ line is, by definition, the Gini metric corresponding with that Lorenz curve. Therefore, by minimizing our Gini metrics, we are equivalently minimizing the areas between each Lorenz curve and its nominal $45°$ line.

E.6. Multi-Period Decomposition Method

Our multi-period ad allocation problem (MG) can be decomposed and solved using repeated invocation of Theorem 1. Here we provide one particular way to achieve a decomposition in this general vein. Computational performance of this particular multi-period decomposition scheme compares favorably with using CPLEX directly. For example, an instance with 4 time periods, 100 campaigns, and 100 viewer types is solved to within 1% of optimality in 12093 seconds using our decomposition method, versus 22483 seconds using CPLEX directly.

Structurally, our proposed multi-period decomposition scheme uses Lagrangian Decomposition to split the problem into a number of single-period-type problems with different definitions for $\hat{\beta}$ and ψ_j, for which we use Theorem 1 to analytically solve all resulting subproblems. To summarize, we break each campaign-j subproblem into two dimensions, a viewer-type dimension and a time dimension. Then, we further decompose the time-dimension subproblem into a series of "viewer-type and time" subproblems. The viewer-type subproblem handles spreading of type (i), the time subproblem handles spreading of type (ii), and the "viewer-type and time" subproblem handles spreading of type (iii). To coordinate the solutions across these subproblems, we use a Dantzig-Wolfe master problem, which now takes convex combinations over both the time-dimension solutions produced at each iteration, and the viewer-type-dimension solutions produced at each iteration. There are additional equality constraints in the Dantzig-Wolfe master problem that ensure that the solutions produced in the time dimension are consistent with those produced in the viewer type dimension; these require additional dual variables and form the basis of the so-called Lagrangian Decomposition. A key structural result is that the "viewer-type and time" subproblem is nested within the time-dimension subproblem; consequently, the solution method uses back-substitution, which turns out to be quite elegant and allows the optimal value of the lower-level "viewer-type and time" subproblem to be analytically embedded into the objective function of the higher-level time-dimension subproblem.

We now describe the details of our decomposition method for our multi-period ad allocation problem (MG). We begin by casting our multi-period ad planning problem (MG) as the following equivalent formulation:

$$
\text{(P-MULT-ORIG)} \qquad \begin{aligned}
\min\ & \alpha_1 \sum_{j\in J} \frac{1}{\hat{s}_j} \sum_{(h,i)\in\Gamma_0(j)} s_h s_i |x_{hj} - x_{ij}| + \alpha_2 \sum_{j\in J} \frac{1}{|T_j|} \sum_{(t,\tau)\in T_0(j)} |w_{jt} - w_{j\tau}| \\
& + \alpha_3 \sum_{j\in J, t\in T_j} \frac{1}{\hat{s}_{jt}} \sum_{(h,i)\in\Gamma_0(j)} s_{ht} s_{it} |x_{hjt} - x_{ijt}| + \sum_{j\in J} p_j y_j \\
\text{s.t.}\ & \sum_{j\in\Gamma(i)\cap J_t} s_{it} x_{ijt} \le s_{it} \qquad \forall i\in I,\ \forall t\in T \qquad \text{(supply)} \\
& \sum_{i\in\Gamma(j), t\in T_j} s_{it} x_{ijt} + y_j = d_j \qquad \forall j\in J \qquad \text{(demand)} \\
& w_{jt} = \sum_{i\in\Gamma(j)} s_{it} x_{ijt} \qquad \forall j\in J,\ \forall t\in T_j \qquad \text{(linking (i))} \\
& s_i x_{ij} = \sum_{t\in T_j} s_{it} x_{ijt} \qquad \forall (i,j)\in\Gamma \qquad \text{(linking (ii))} \\
& x_{ij} \ge 0\ \forall (i,j)\in\Gamma; \qquad x_{ijt} \ge 0\ \forall j\in J, \forall i\in\Gamma(j), \forall t\in T_j \\
& w_{jt} \ge 0\ \forall j\in J,\ \forall t\in T_j; \qquad y_j \ge 0\ \forall j\in J
\end{aligned}
$$

Noticing that there are two dimensions to the planning problem, the viewer-type dimension and the time dimension, we exploit this structure by not only designing a decomposition scheme that decomposes the problem by campaign, but also producing a scheme that further decomposes each campaign's subproblem into a viewer-type subproblem and a time subproblem. Additionally, since we want to leverage the results provided by Theorem 1, we need both the viewer-type subproblem for campaign j and the time subproblem for campaign j to each have its own set of decision variables and its own demand constraint. Therefore, we introduce the decision variables y_j^T, $j\in J$, which we initially fix to be equal to their respective y_j, $j\in J$, counterparts. The y_j variables will measure demand shortfalls in the viewer-type subproblems, while y_j^T will measure demand shortfalls in the time subproblems (the T superscript denotes the time dimension). While it is the case that $y_j = y_j^T$ for all campaigns $j \in J$ must hold for any feasible plan, it could be the case that at any iteration, the subproblems produce solutions that have $y_j \neq y_j^T$. This can and often does happen because the viewer-type subproblems are solved independently from the time subproblems. We will describe later how we use Lagrangian Decomposition to produce solutions that get close to satisfying $y_j = y_j^T\ \forall j \in J$; for now, it is sufficient to understand that while we would like y_j and y_j^T to take on the same values, we model these variables separately so that we can decouple the campaign-j subproblems into viewer-type and time dimensions. Thus, we produce two logically equivalent sets of demand constraints, one using the y_j variables which will end up in the viewer-type subproblems, and the other using the y_j^T variables which will end up in the time subproblems. With these additions to our formulation, (P-MULT-ORIG) can be equivalently represented as:

$$
\text{(P-MULT-ORIG2)} \qquad \min\ \alpha_1 \sum_{j\in J} \frac{1}{\hat{s}_j} \sum_{(h,i)\in\Gamma_0(j)} s_h s_i |x_{hj} - x_{ij}| + \alpha_2 \sum_{j\in J} \frac{1}{|T_j|} \sum_{(t,\tau)\in T_0(j)} |w_{jt} - w_{j\tau}|
$$
$$
+ \alpha_3 \sum_{j\in J, t\in T_j} \frac{1}{\hat{s}_{jt}} \sum_{(h,i)\in\Gamma_0(j)} s_{ht} s_{it} |x_{hjt} - x_{ijt}| + \frac{1}{2}\sum_{j\in J} p_j (y_j + y_j^T)
$$
$$
\begin{aligned}
\text{s.t.} \sum_{j\in\Gamma(i)\cap J_t} s_{it} x_{ijt} &\le s_{it} && \forall i\in I,\ \forall t\in T && \text{(supply)}\\
\sum_{t\in T_j} w_{jt} + y_j^T &= d_j && \forall j\in J && \text{(demand (i))}\\
\sum_{i\in\Gamma(j)} s_i x_{ij} + y_j &= d_j && \forall j\in J && \text{(demand (ii))}\\
w_{jt} &= \sum_{i\in\Gamma(j)} s_{it} x_{ijt} && \forall j,\ \forall t\in T_j && \text{(linking (i))}\\
s_i x_{ij} &= \sum_{t\in T_j} s_{it} x_{ijt} && \forall (i,j)\in\Gamma && \text{(linking (ii))}\\
y_j &= y_j^T && \forall j\in J && \text{(linking (iii))}\\
& x_{ij} \ge 0\ \forall (i,j)\in\Gamma; \qquad x_{ijt}\ge 0\ \forall j\in J, \forall i\in\Gamma(j), \forall t\in T_j\\
& w_{jt}\ge 0\ \forall j\in J,\ \forall t\in T_j; \qquad y_j\ge 0\ \forall j\in J; \qquad y_j^T \ge 0\ \forall j\in J
\end{aligned}
$$

We now produce a relaxation of (P-MULT-ORIG2) which dualizes the supply constraint, as well as linking constraints (ii) and (iii). We use dual variables $\hat{\beta}_{it}$ for the supply constraints, as well as dual variables η_{ij} and γ_j for linking constraints (ii) and (iii), respectively. Note that γ_j and η_{ij} can take negative or positive values, whereas $\hat{\beta}_{it} \ge 0$. The relaxed problem is:

$$
\text{(PS-MULT)} \quad z^{LB} = \min\ \alpha_1 \sum_{j\in J} \frac{1}{\hat{s}_j} \sum_{(h,i)\in\Gamma_0(j)} s_h s_i |x_{hj} - x_{ij}| + \alpha_2 \sum_{j\in J} \frac{1}{|T_j|} \sum_{(t,\tau)\in T_0(j)} |w_{jt} - w_{j\tau}|
$$
$$
+ \alpha_3 \sum_{j\in J, t\in T_j} \frac{1}{\hat{s}_{jt}} \sum_{(h,i)\in\Gamma_0(j)} s_{ht} s_{it} |x_{hjt} - x_{ijt}| + \frac{1}{2}\sum_{j\in J} p_j (y_j + y_j^T) + \sum_{j\in J} \gamma_j (y_j - y_j^T)
$$

$$
\begin{aligned}
&+ \sum_{(i,j)\in\Gamma} \eta_{ij} \left(s_i x_{ij} - \sum_{t\in T_j} s_{it} x_{ijt} \right) + \sum_{i\in I, t\in T} \hat{\beta}_{it} \left(\sum_{j\in\Gamma(i)\cap J_t} s_{it} x_{ijt} - s_{it} \right) \\
\text{s.t.} &\sum_{t\in T_j} w_{jt} + y_j^T = d_j \qquad \forall j \in J \qquad \text{(demand (i))} \\
&\sum_{i\in\Gamma(j)} s_i x_{ij} + y_j = d_j \quad \forall j \in J \qquad \text{(demand (ii))} \\
&w_{jt} = \sum_{i\in\Gamma(j)} s_{it} x_{ijt} \qquad \forall j,\, \forall t \in T_j \quad \text{(linking (i))} \\
&x_{ij} \geq 0\, \forall (i,j) \in \Gamma; \qquad x_{ijt} \geq 0\, \forall j \in J, \forall i \in \Gamma(j), \forall t \in T_j \\
&w_{jt} \geq 0\, \forall j \in J,\, \forall t \in T_j; \qquad y_j \geq 0\, \forall j \in J; \qquad y_j^T \geq 0\, \forall j \in J
\end{aligned}
$$

We can also simplify the objective function further, to get:

$$
\begin{aligned}
\min\ &\alpha_1 \sum_{j\in J} \frac{1}{\hat{s}_j} \sum_{(h,i)\in\Gamma_0(j)} s_h s_i |x_{hj} - x_{ij}| + \alpha_2 \sum_{j\in J} \frac{1}{|T_j|} \sum_{(t,\tau)\in T_0(j)} |w_{jt} - w_{j\tau}| \\
&+ \alpha_3 \sum_{j\in J, t\in T_j} \frac{1}{\hat{s}_{jt}} \sum_{(h,i)\in\Gamma_0(j)} s_{ht} s_{it} |x_{hjt} - x_{ijt}| + \sum_{j\in J} \left(\frac{1}{2} p_j + \gamma_j \right) y_j \\
&+ \sum_{j\in J} (\frac{1}{2} p_j - \gamma_j) y_j^T + \sum_{j\in J, i\in\Gamma(j)} \eta_{ij} s_i x_{ij} + \sum_{j\in J, i\in\Gamma(j), t\in T_j} (\hat{\beta}_{it} - \eta_{ij}) s_{it} x_{ijt} - \sum_{i\in I, t\in T} \hat{\beta}_{it} s_{it}
\end{aligned}
$$

With supply, linking (ii), and linking (iii) constraints dualized, we can see that the relaxed problem (PS-MULT) decomposes into one viewer-type subproblem and one time subproblem for each campaign. The viewer-type subproblem for campaign j has decision variables $\{x_{ij}\}_{(i,j)\in\Gamma}$ and y_j, has optimal value $z_j^{VLB}(\hat{\beta}, \eta, \gamma)$, and is defined as:

$$
\begin{aligned}
\text{(PS-V-j)} \qquad z_j^{VLB}(\hat{\beta}, \eta, \gamma) = \min\ &\alpha_1 \frac{1}{\hat{s}_j} \sum_{(h,i)\in\Gamma_0(j)} s_h s_i |x_{hj} - x_{ij}| + \left(\frac{1}{2} p_j + \gamma_j \right) y_j + \sum_{i\in\Gamma(j)} \eta_{ij} s_i x_{ij} \\
\text{s.t.} &\sum_{i\in\Gamma(j)} s_i x_{ij} + y_j = d_j \qquad \text{(demand (ii))} \\
&x_{ij} \geq 0\, \forall i \in \Gamma(j); \qquad y_j \geq 0
\end{aligned}
$$

Similarly, the time subproblem for campaign j has decision variables $\{x_{ijt}\}_{i\in\Gamma(j),\forall t\in T_j}$, $\{w_{jt}\}_{t\in T_j}$, and y_j^T, has optimal value $z_j^{TLB}(\hat{\beta}, \eta, \gamma)$, and is defined as:

$$
\begin{aligned}
\text{(PS-T-j)} \qquad z_j^{TLB}(\hat{\beta}, \eta, \gamma) = \min\ &\alpha_2 \frac{1}{|T_j|} \sum_{(t,\tau)\in T_0(j)} |w_{jt} - w_{j\tau}| + \alpha_3 \sum_{t\in T_j} \frac{1}{\hat{s}_{jt}} \sum_{(h,i)\in\Gamma_0(j)} s_{ht} s_{it} |x_{hjt} - x_{ijt}| \\
&+ \left(\frac{1}{2} p_j - \gamma_j \right) y_j^T + \sum_{i\in\Gamma(j), t\in T_j} (\hat{\beta}_{it} - \eta_{ij}) s_{it} x_{ijt} \\
\text{s.t.} &\sum_{t\in T_j} w_{jt} + y_j^T = d_j \qquad \text{(demand (i))} \\
&w_{jt} = \sum_{i\in\Gamma(j)} s_{it} x_{ijt} \quad \forall t \in T_j \quad \text{(linking (i))} \\
&x_{ijt} \geq 0\, \forall i \in \Gamma(j),\, \forall t \in T_j; \qquad w_{jt} \geq 0\, \forall t \in T_j; \qquad y_j^T \geq 0
\end{aligned}
$$

The relaxed problem (PS-MULT) is related to the set of $|J|$ subproblems (PS-V-j) and (PS-T-j) via the following formula, which expresses the optimal value of (PS-MULT) in terms of the optimal values of (PS-V-j) and (PS-T-j):

$$z^{LB} = -\sum_{i \in I, t \in T} \hat{\beta}_{it} s_{it} + \sum_{j \in J} z_j^{VLB} + \sum_{j \in J} z_j^{TLB}$$

Theorem 1 tells us how to solve (PS-V-j), the viewer-type subproblem for campaign j. We take $\psi_j^V = \alpha_1 / \hat{s}_j$, treat $p_j^V = (\frac{1}{2} p_j + \gamma_j)$ as the p_j defined in the theorem, and treat $\hat{\beta}_i^{Vj} = \eta_{ij}$ as the $\hat{\beta}_i$ defined in the theorem. Invoking the theorem yields the following. Assuming without loss of generality that all $i \in \Gamma(j)$ viewer types are sorted by $\hat{\beta}_i^{Vj}$, i.e., $\hat{\beta}_1^{Vj} \le \hat{\beta}_2^{Vj} \le \cdots \le \hat{\beta}_{m_j}^{Vj}$, where $m_j = |\Gamma(j)|$, we define $s_i^{Vbj} = \sum_{i'=1..i-1} s_{i'}$; $s_i^{Vaj} = \sum_{i'=i+1..m_j} s_{i'}$; $c_{ij}^V = \psi_j^V s_i (s_i^{Vaj} - s_i^{Vbj}) + s_i \hat{\beta}_i^{Vj}$; $\tilde{c}_{ij}^V = \sum_{i'=1..i} c_{i'j}^V$; $\tilde{s}_{ij}^V = \sum_{i'=1..i} s_{i'}$; $\pi_{ij}^V = \tilde{c}_{ij}^V / \tilde{s}_{ij}^V$; and $i^* = \arg\min_{i \in \{1..m_j\}} \pi_{ij}^V$. If $\pi_{i^*j}^V > p_j^V$, then $y_j^* = d_j$, $x_{ij}^* = 0 \; \forall i = 1..m_j$, and the corresponding optimal value is $p_j^V d_j$. Otherwise, the optimal value is $\pi_{i^*j}^V d_j$ with corresponding optimal solution $y_j^* = 0$, and

$$x_{ij}^* = \begin{cases} d_j / \tilde{s}_{i^*j}^V & \text{for } i \le i^* \\ 0 & \text{for } i > i^* \end{cases}$$

Solving (PS-T-j), the time subproblem for campaign j, is more involved. For this, we make use of a very nice further decomposition. Indeed, we break (PS-T-j) into a number of "viewer-type and time" subproblems, each of which is responsible for spreading impressions across viewer types for a single time period. Each such "viewer-type and time" subproblem has the required structure to invoke Theorem 1, and thus can be solved analytically. Moreover, via backwards substitution, we can encode the optimal values from these "viewer-type and time" subproblems into the objective function of the time subproblem, finally solve the time subproblem via another invocation of Theorem 1, and then analytically compute the optimal solution to each "viewer-type and time" subproblem given the optimal solution to the time subproblem. To derive this construction, we first define (PS-VT-jt), the campaign-j subproblem which spreads impressions across viewer types within a single time period t. The subproblem (PS-VT-jt) has decision variables $\{x_{ijt}\}_{i \in \Gamma(j) \cap T_j}$, has optimal value $z_{jt}^{VTLB}(\hat{\beta}, \eta, w_{jt})$, and is defined as:

$$\text{(PS-VT-jt)} \qquad z_{jt}^{VTLB}(\hat{\beta}, \eta, w_{jt}) = \min \; \alpha_3 \sum_{t \in T_j} \frac{1}{\hat{s}_{jt}} \sum_{(h,i) \in \Gamma_0(j)} s_{ht} s_{it} |x_{hjt} - x_{ijt}| + \sum_{i \in \Gamma(j), t \in T_j} (\hat{\beta}_{it} - \eta_{ij}) s_{it} x_{ijt}$$

$$\text{s.t.} \sum_{i \in \Gamma(j)} s_{it} x_{ijt} = w_{jt} \quad \text{(linking (i))}$$

$$x_{ijt} \ge 0 \; \forall i \in \Gamma(j) \cap T_j$$

Using the optimal value from (PS-VT-jt), we can write the campaign-j time subproblem (PS-T-j) equivalently as:

$$\text{(PS-T2-j)} \qquad z_j^{TLB}(\hat{\beta}, \eta, \gamma) = \min \; \alpha_2 \frac{1}{|T_j|} \sum_{(t,\tau) \in T_0(j)} |w_{jt} - w_{j\tau}| + (\frac{1}{2} p_j - \gamma_j) y_j^T + \sum_{t \in T_j} z_{jt}^{VTLB}(\hat{\beta}, \eta, w_{jt})$$

$$\text{s.t.} \sum_{t \in T_j} w_{jt} + y_j^T = d_j \quad \text{(demand (i))}$$

$$w_{jt} \ge 0 \; \forall t \in T_j; \qquad y_j^T \ge 0$$

With this representation, (PS-T2-j) encodes $|T_j|$ inner optimization problems of the form (PS-VT-jt) into its objective function.

Theorem 1 tells us how to solve (PS-VT-jt), the campaign-j subproblem for spreading impressions across viewer types within time period t. We take $\psi_j^t = \alpha_3/\hat{s}_{jt}$, treat $\hat{\beta}_i^{tj} = (\hat{\beta}_{it} - \eta_{ij})$ as the $\hat{\beta}_i$ defined by the theorem, and treat w_{jt} as d_j. Moreover, since no shortfall is allowed for the "demand" constraint $\sum_{i\in\Gamma(j)} s_{it}x_{ijt} = w_{jt}$, we treat p_j as defined by the theorem as if it is $+\infty$ so that no shortfall occurs in the solution. Invoking the theorem yields the following. Assuming without loss of generality that all $i \in \Gamma(j)\cap T_j$ viewer types are sorted by $\hat{\beta}_i^{tj}$, i.e., $\hat{\beta}_1^{tj} \le \hat{\beta}_2^{tj} \le \cdots \le \hat{\beta}_{m_j}^{tj}$, where $m_j = |\Gamma(j)\cap T_j|$, we define $s_i^{tbj} = \sum_{i'=1..i-1} s_{i't}$; $s_i^{taj} = \sum_{i'=i+1..m_j} s_{i't}$; $c_{ij}^t = \psi_j^t s_{it}(s_i^{taj} - s_i^{tbj}) + s_{it}\hat{\beta}_i^{tj}$; $\tilde{c}_{ij}^t = \sum_{i'=1..i} c_{i'j}^t$; $\tilde{s}_{ij}^t = \sum_{i'=1..i} s_{i't}$; $\pi_{ij}^t = \tilde{c}_{ij}^t/\tilde{s}_{ij}^t$; and $i^* = \arg\min_{i\in\{1..m_j\}} \pi_{ij}^t$. Then, the optimal value of (PS-VT-jt) corresponding to campaign-j and time period t is $\pi_{i^*j}^t w_{jt}$, with corresponding optimal solution:

$$x_{ijt}^* = \begin{cases} w_{jt}/\tilde{s}_{i^*j}^t & \text{for } i \le i^* \\ 0 & \text{for } i > i^* \end{cases} \tag{EC.11}$$

Moreover, denoting $\pi_{jt}^* \equiv \pi_{i^*j}^t$, we have $z_{jt}^{VTLB}(\hat{\beta}, \eta, w_{jt}) = \pi_{jt}^* w_{jt}$, where Theorem 1 gives us π_{jt}^* as a function of $\hat{\beta}$ and η. Thus, the campaign-j time subproblem (PS-T2-j) can be simplified to:

$$\text{(PS-T3-j)} \qquad z_j^{TLB}(\hat{\beta}, \eta, \gamma) = \min\ \alpha_2 \frac{1}{|T_j|} \sum_{(t,\tau)\in T_0(j)} |w_{jt} - w_{j\tau}| + \left(\frac{1}{2}p_j - \gamma_j\right) y_j^T + \sum_{t\in T_j} \pi_{jt}^* w_{jt}$$
$$\text{s.t.} \sum_{t\in T_j} w_{jt} + y_j^T = d_j \qquad \text{(demand (i))}$$
$$w_{jt} \ge 0\ \forall t \in T_j; \qquad y_j^T \ge 0$$

Now we notice that the time subproblem (PS-T3-j) is in the form required to also solve it using Theorem 1. We take $\psi_j^T = \alpha_2/|T_j|$, treat $p_j^T = (\frac{1}{2}p_j - \gamma_j)$ as the p_j defined in the theorem, and treat $\hat{\beta}_t^{Tj} = \pi_{jt}^*$ as the $\hat{\beta}_i$ defined in the theorem. Moreover, w_{jt} takes the place of x_{ij} and correspondingly we assume $s_i = 1$ for all viewer types $i \in I$. Invoking the theorem yields the following. Assuming without loss of generality that all $t \in T_j$ time periods are sorted by $\hat{\beta}_t^{Tj}$, i.e., $\hat{\beta}_1^{Tj} \le \hat{\beta}_2^{Tj} \le \cdots \le \hat{\beta}_{m_j}^{Tj}$, where $m_j = |T_j|$, we define $s_t^{Tbj} = \sum_{t'=1..t-1} 1 = t-1$; $s_t^{Taj} = \sum_{t'=t+1..m_j} 1 = m_j - t$; $c_{tj}^T = \psi_j^T(s_t^{Taj} - s_t^{Tbj}) + \hat{\beta}_t^{Tj}$; $\tilde{c}_{tj}^T = \sum_{t'=1..t} c_{t'j}^T$; $\tilde{s}_{tj}^T = \sum_{t'=1..t} 1 = t$; $\pi_{tj}^T = \tilde{c}_{tj}^T/\tilde{s}_{tj}^T$; and $t^* = \arg\min_{t\in\{1..m_j\}} \pi_{tj}^T$. If $\pi_{t^*j}^T > p_j^T$, then $y_j^{T*} = d_j$, $w_{jt}^* = 0\ \forall t = 1..m_j$, and the corresponding optimal value is $p_j^T d_j$. Otherwise, the optimal value is $\pi_{t^*j}^T d_j$ with corresponding optimal solution $y_j^{T*} = 0$, and

$$w_{jt}^* = \begin{cases} d_j/t^* & \text{for } t \le t^* \\ 0 & \text{for } t > t^* \end{cases} \tag{EC.12}$$

Finally, once we have solution (EC.12) for (PS-T-j), we can substitute it into (EC.11) to yield the solutions for (PS-VT-jt) corresponding to all $t \in T_j$.

We have now shown how to completely solve all subproblems analytically, which can be summed up as, for each campaign j, (1) solving (PS-V-j) analytically using Theorem 1; (2) solving (PS-VT-jt) analytically using Theorem 1 for all time periods $t \in T_j$; (3) using the optimal values π_{jt}^* computed for (PS-VT-jt), $t \in T_j$, to solve

(PS-T-j) analytically using Theorem 1; and then finally (4) numerically computing the solutions to (PS-VT-jt), $t \in T_j$, using the optimal solution from (PS-T-j) by substituting (EC.12) into (EC.11). Equivalently, we are able to solve (PS-MULT) analytically.

There are two major elements of our decomposition method that still need to be developed before we can solve the original problem, (P-MULT-ORIG). As in the single-period decomposition scheme discussed in Section 6 and Appendix C, we additionally need (1) a method for choosing a "good" price vector $(\hat{\beta}, \eta, \gamma)$; and (2) a method to convert one or more solutions to (PS-MULT) into a near-optimal solution for (P-MULT-ORIG) that satisfies all of the constraints that were dualized when relaxing (P-MULT-ORIG) to produce (PS-MULT) (i.e., the supply constraints, as well as linking constraints (ii) and (iii)). We proceed as in the single-period case, and develop a Dantzig-Wolfe decomposition scheme that we modify for performance reasons to solve the Dantzig-Wolfe master problem only once every five iterations, instead updating the price vector $(\hat{\beta}, \eta, \gamma)$ using subgradient optimization in the intervening four of five iterations.

Our solution method begins by first solving a heuristic to get an initial feasible solution. Our heuristic has two stages. In stage one, the heuristic iterates through the list of campaigns $j = 1..|J|$, and attempts to assign $s_{it}\theta_j$ impressions of each viewer type $i \in \Gamma(j)$ in each time period $t \in T_j$ to campaign j. Such allocations are processed in sequence, and at some point the remaining supply of viewer type i for time period t may be less than the $s_{it}\theta_j$ impressions campaign j is requesting. At such point, the campaign j gets the remaining supply of s_{it}, and subsequent campaigns do not get any impressions from viewer type i at time t. When stage one is completed, some viewer types in some time periods may be completely allocated, and yet some campaigns may still have demand that has not been allocated. This can happen whenever a campaign cannot grab all $s_{it}\theta_j$ impressions from each viewer type $i \in \Gamma(j)$ and time period $t \in T_j$ that it wants. Stage two then iterates through the list of campaigns once again, but this time each campaign j checks to see how many impressions in each matching viewer type $i \in \Gamma(j)$ and time period $t \in T_j$ are still available, and grabs the same proportion of slack from each matching viewer type in each time period. Ideally, after booking this proportion of slack, the demand for campaign j is fully allocated. However, it could be that the proportion of slack needed to satisfy demand would exceed one; in that case, the campaign grabs all remaining slack from each matching viewer type in each matching time period, and remains partially unallocated. We denote the solution computed by this heuristic as $\left\{\{x_{ijt0}\}_{(i,j)\in\Gamma}, \{y_{j0}\}_{j\in J}\right\}$, and compute the values corresponding to each j-component $\{v_{j0}\}_{j\in J}$ using the terms corresponding to campaign j in the Gini-based objective (EC.9). In other words, for $n = 0$, we use linking constraints (i), (ii), and (iii) to first compute $x_{ijn} = (1/s_i)\sum_{t\in T_j} s_{it}x_{ijtn}$; $w_{jtn} = \sum_{i\in\Gamma(j)} s_{it}x_{ijtn}$; and $y^T_{jn} = y_{jn}$; and then we substitute those values into the following formulas:

$$v^V_{jn} := \alpha_1 \frac{1}{\hat{s}_j} \sum_{(h,i)\in\Gamma_0(j)} s_h s_i |x_{hjn} - x_{ijn}| + \frac{1}{2} p_j y_{jn}, \text{ and} \tag{EC.13}$$

$$v^T_{jn} := \alpha_2 \frac{1}{|T_j|} \sum_{(t,\tau)\in T_0(j)} |w_{jtn} - w_{j\tau n}| + \alpha_3 \sum_{t\in T_j} \frac{1}{\hat{s}_{jt}} \sum_{(h,i)\in\Gamma_0(j)} s_{ht}s_{it}|x_{hjtn} - x_{ijtn}| + \frac{1}{2} p_j y^T_{jn}. \tag{EC.14}$$

We then use this heuristic solution to initialize the first set of columns, corresponding to $n=0$, in the Dantzig-Wolfe master problem. The Dantzig-Wolfe master problem, which has parameters x_{ijtn}, x_{ijn}, y_{jn}, y^T_{jn}, v^T_{jn}, and v^V_{jn}, $n=0..N$, $j\in J$, $i\in\Gamma(j)$, $t\in T_j$, and decision variables λ^V_{jn}, λ^T_{jn}, $n=0..N$, $j\in J$, is formulated as follows:

$$
\begin{aligned}
&\text{(PM-MULT)} \qquad z^{UB}=\min \sum_{j\in J,n=0..N}(v^V_{jn}\lambda^V_{jn}+v^T_{jn}\lambda^T_{jn}) && \\
&\text{s.t.} \sum_{n=0..N.j\in\Gamma(it)} x_{ijtn}\lambda^T_{jn}\le 1 && \forall i\in I,\ \forall t\in T \quad \text{(supply)}\\
&\sum_{n=0..N} s_i x_{ijn}\lambda^V_{jn}=\sum_{n=0..N}\sum_{t\in T_j} s_{it}x_{ijtn}\lambda^T_{jn} && \forall (i,j)\in\Gamma \quad \text{(linking (ii))}\\
&\sum_{n=0..N} y_{jn}\lambda^V_{jn}=\sum_{n=0..N} y^T_{jn}\lambda^T_{jn} && \forall j\in J \quad \text{(linking (iii))}\\
&\sum_{n=0..N}\lambda^V_{jn}=1 && \forall j\in J \quad \text{(convexity)}\\
&\sum_{n=0..N}\lambda^T_{jn}=1 && \forall j\in J \quad \text{(convexity)}\\
&\lambda^V_{jn}\ge 0,\ \lambda^T_{jn}\ge 0 && \forall j\in J,\ \forall n=0..N
\end{aligned}
$$

In a classical Dantzig-Wolfe decomposition, we would begin with $N=0$; solve the master problem (PM-MULT); take $\hat{\beta}_{it}$ to be the dual values of the supply constraints for each viewer type $i\in I$ at each time period $t\in T$; take η_{ij} to be the dual values of the “linking (ii)” constraints for each $(i,j)\in\Gamma$; take γ_j to be the dual values of the “linking (iii)” constraints for each $j\in J$; re-solve the subproblems (PS-V-j), (PS-VT-jt), and (PS-T-j) for all campaigns $j\in J$ and time periods $t\in T_j$ (analytically, via repeated use of Theorem 1, as we have described); increment the iteration counter $N:=N+1$ and record the subproblem solutions as $\{x_{ijtN}\}_{j\in J,i\in\Gamma(j),t\in T_j}$, $\{x_{ijN}\}_{j\in J,i\in\Gamma(j)}$, $\{y_{jN}\}_{j\in J}$, $\{y^T_{jN}\}_{j\in J}$, with values $\{v^V_{jN}\}_{j\in J}$ and $\{v^T_{jN}\}_{j\in J}$, computed again using (EC.13) and (EC.14); add the subproblem solutions as a new set of columns to the master problem; and then iterate, repeatedly solving the master problem (PM-MULT) and subproblems (PS-V-j), (PS-VT-jt), and (PS-T-j) for all $j=1..|J|$, $t\in T_j$, in this fashion until a termination criterion is attained.

We use the master problem (PM-MULT) to produce a near-optimal (and feasible) solution to (P-MULT-ORIG), where the variables λ^V_{jn} and λ^T_{jn} determine the weights to apply to the solutions produced by the viewer-type and time subproblems for campaign j in iteration n, respectively. The master problem (PM-MULT) yields an upper bound and the subproblems collectively yield the lower bound $z^{LB}(\hat{\beta},\eta,\gamma)$. Therefore, at each iteration, we can compute the optimality gap $(z^{UB}-z^{LB})/z^{LB}$, and terminate when it is below a desired threshold.

As in the single-period case, we solve the master problem (PM) only every fifth iteration, and employ subgradient optimization to update the prices $\hat{\beta}_{it}$, $i\in I$, $t\in T$; η_{ij}, $(i,j)\in\Gamma$; γ_j, $j\in J$ for four out of every five iterations. After we update the price vector using subgradient optimization, we solve the subproblems (PS-V-j), (PS-VT-jt), and (PS-T-j) for all campaigns $j\in J$ and time periods $t\in T_j$, just as before, and update the master problem's columns using the subproblem solutions.

To use subgradient optimization, we note that the $(i,t)^{th}$ component of the subgradient corresponding to the $(i,t)^{th}$ relaxed supply constraint is $g_{it} := \sum_{j\in\Gamma(i)\cap J_t} s_{it}x_{ijt} - s_{it}$; the $(i,j)^{th}$ component of the subgradient corresponding to the $(i,j)^{th}$ relaxed "linking (ii)" constraint is $g_{ij} := s_i x_{ij} - \sum_{t\in T_j} s_{it}x_{ijt}$; and the j^{th} component of the subgradient corresponding to the j^{th} relaxed "linking (iii)" constraint is $g_j := y_j - y_j^T$. Using the step size σ_n at iteration n, we update the price vector $(\hat{\beta},\eta,\gamma)$ using subgradient optimization as follows:

$$\hat{\beta}_{it} := \max(0, \hat{\beta}_{it} + \sigma_n g_{it});$$

$$\eta_{ij} := \eta_{ij} + \sigma_n g_{ij};$$

$$\gamma_j := \gamma_j + \sigma_n g_j.$$

We use step sizes

$$\sigma_n = \frac{2(z^{UB} - z^{LB}(\hat{\beta},\eta,\gamma))}{\sum_{i\in I,t\in T} g_i^2 + \sum_{(i,j)\in\Gamma} g_{ij}^2 + \sum_{j\in J} g_j^2},$$

which have been shown in the literature (e.g., Fisher 2004) to make steady progress toward the optimal dual prices.

To summarize, we solve (P-MULT-ORIG) to near-optimality by (1) initializing the master problem (PM-MULT) at iteration $n=0$ using a heuristic solution; (2) incrementing the iteration counter $(N := N + 1)$; (3a) solving the master problem (PM-MULT) every fifth iteration to get a new price vector $(\hat{\beta},\eta,\gamma)$, a near-optimal feasible solution, and the value of this solution which serves as an upper bound z^{UB} for (P-MULT-ORIG); (3b) employing subgradient optimization for four out of every five iterations to update the values of $(\hat{\beta},\eta,\gamma)$, from the past iteration; (4) solving all campaign-j subproblems for the viewer-type, time, and "viewer-type and time" dimensions analytically using Theorem 1 to get new columns to be placed into the master problem (PM-MULT) as well as a new lower bound z^{LB}; (5) updating either the lower bound or the upper bound, or both, if they are tighter than they were last iteration; (6) computing the optimality gap $(z^{UB} - z^{LB})/z^{LB}$ and terminating if the termination criterion has been attained (e.g., optimality gap $< 1\%$); or continuing to iterate by going to step 2, otherwise.

E.7. Tail Gini

In the baseline multi-period Gini model (MB) from §E.4, penalties are incurred if the impression allocation does not fall within the impression target intervals $[l_{ij}, u_{ij}]$, $[l_{jt}, u_{jt}]$, and $[l_{ijt}, u_{ijt}]$ defined by the constraints (EC.5d), (EC.5e), and (EC.5f). To formulate Gini-based models that incorporate the idea of impression target ranges, where no penalty is incurred when the impression allocation is within the target range, we use the tail-Gini concept (see the related studies of Mansini et al. 2007; Ogryczak and Ruszczyński 2002a,b). Consider an n-dimensional vector x. The tail-Gini coefficient of x is the conditional average of the absolute differences between the components x_i and x_j of each pair $(x_i, x_j), i,j = 1,\ldots,n$ divided by twice the mean value $\mu = \frac{1}{n}\sum_{i=1}^n x_i$:

$$TG = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} (|x_i - x_j| - r)^+}{2n^2\mu},$$

where r is a user-defined threshold value. The above formula shows that the tail-Gini measure is based on the idea of conditional deviation and is equal to the standard Gini (1) if r is equal to 0. A pair (x_i, x_j) has an impact on the value of TG if and only if the absolute value of the difference between x_i and x_j is strictly larger than r.

Using the above definition, we may formulate tail-Gini metrics corresponding to (5), (EC.6), and (EC.7) as:

$$TG_j = \left(\frac{1}{\hat{s}_j}\right) \frac{\sum\limits_{(h,i)\in\Gamma_0(j)} s_h s_i \left(|x_{hj} - x_{ij}| - r_{hi}\right)^+}{\sum\limits_{i\in\Gamma(j)} s_i x_{ij}} \ ,$$

$$TG_{jt} = \left(\frac{1}{\hat{s}_{jt}}\right) \frac{\sum\limits_{(h,i)\in\Gamma_0(j)} s_{ht} s_{it} \left(|x_{hjt} - x_{ijt}| - r_{hit}\right)^+}{\sum_{i\in\Gamma(j)} s_{it} x_{ijt}} \ ,$$

and

$$TG_j^T = \left(\frac{1}{|T_j| w_j}\right) \sum_{(t,\tau)\in T_0(j)} \left(|w_{jt} - w_{j\tau}| - r_t\right)^+ \ ,$$

where r_{hi}, r_{hit}, and r_t are fixed parameters.

Using the tail-Gini metric, we obtain the following objective function:

$$\min \ \alpha_1 \sum_{j\in J} w_j TG_j + \alpha_2 \sum_{j\in J} w_j TG_j^T + \alpha_3 \sum_{j\in J, t\in T_j} w_{jt} TG_{jt} + \sum_{j\in J} p_j y_j \ .$$

References

Araman V, Fridgeirsdottir K (2015) Cost-per-impression pricing and campaign delivery for online display advertising. Working Paper, American University of Beirut, Beirut, Lebanon. Available at `http://pages.stern.nyu.edu/~varaman/Research/OnlineAd.pdf`.

Bhalgat V, Feldman J, Mirrokni V (2012) Online allocation of display ads with smooth delivery. In *Proc. 18th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Beijing, China, pp. 1213–1221.

Bollapragada S, Cheng H, Phillips M, Garbiras M, Scholes M, Gibbs T, Humphreville M (2002) NBC's optimization systems increase its revenues and productivity. *Interfaces* 32(1): 47–60.

Fisher M (2004) The Lagrangian relaxation method for solving integer programming problems. *Management Science* 50(12): 1861–1871.

Lee KC, Jalali A, Dasdan A (2013) Real time bid optimization with smooth budget delivery in online advertising. In *Proc. 7th International Workshop on Data Mining for Online Advertising*, Chicago, Illinois, Article 1.

Mansini R, Ogryczak W, Speranza MG (2007) Tail Gini's risk measures and related linear programming models for portfolio optimization. In *HERCMA Conference Proceedings, CD, LEA Publishers, Athens*. Available at `http://www.ia.pw.edu.pl/~wogrycza/publikacje/artykuly/mcvarat.pdf`.

Ogryczak W, Ruszczyński A (2002a) Dual stochastic dominance and related mean-risk models. *SIAM Journal on Optimization* 13: 60–78.

Ogryczak W, Ruszczyński A (2002b) Dual stochastic dominance and quantile risk measures. *International Transactions in Operational Research* 9: 661–680.

Turner J, Scheller-Wolf A, Tayur S (2011) Scheduling of dynamic in-game advertising. *Operations Research* 59(1): 1–16.

Xu J, Lee KC, Li W, Qi H, Lu Q (2015) Smart pacing for effective online ad campaign optimization. In *Proc. 21st ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Sydney, NSW, Australia, pp. 2217–2226.