Supplementary Appendix: Causal Inference for Aggregated Treatment

Carolina Caetano[†] Gregorio Caetano[‡] Brantly Callaway[§] Derek Dyal[¶]

December 12, 2025

The Supplementary Appendix contains proofs of supporting results from the main text, along with additional details and results for the application about enrichment activities.

SA Supplementary Proofs

SA.1 Proofs of Results from Section 3

Proof of Proposition C.2. To show the first part, notice that, under Assumption 1,

$$\begin{aligned}\mathbb{E}[Y|S=s_d]-\mathbb{E}[Y|S=s_d'] &= \mathbb{E}[Y(s_d)|S=s_d]-\mathbb{E}[Y(s_d')|S=s_d'] \\ &= \mathbb{E}[Y(s_d)-Y(s_d')|S=s_d]+\mathbb{E}[Y(s_d')|S=s_d]-\mathbb{E}[Y(s_d')|S=s_d'] \\ &= \text{SATT}(s_d,s_d')+SB(s_d,s_d'),\end{aligned}$$

where the first equality holds by writing observed outcomes in terms of their corresponding potential outcomes; the second equality holds by adding and subtracting $\mathbb{E}[Y(s_d')|S=s_d]$; and the third equality holds by the definitions of $\text{SATT}(s_d,s_d')$ and $SB(s_d,s_d')$.

For the second part, suppose that Assumption 6 holds. Notice that

$$\begin{aligned}\text{SATT}(s_d,s_d') &= \mathbb{E}[Y(s_d)-Y(s_d')|S=s_d] \\ &= \mathbb{E}[Y(s_d)|S=s_d]-\mathbb{E}[Y(s_d')|S=s_d] \\ &= \mathbb{E}[Y(s_d)|S=s_d]-\mathbb{E}[Y(s_d')|S=s_d,D=d] \\ &= \mathbb{E}[Y(s_d)|S=s_d]-\mathbb{E}[Y(s_d')|S=s_d',D=d] \\ &= \mathbb{E}[Y(s_d)|S=s_d]-\mathbb{E}[Y(s_d')|S=s_d'] \\ &= \mathbb{E}[Y|S=s_d]-\mathbb{E}[Y|S=s_d'],\end{aligned}$$

where the first equality hold by the definition of $\text{SATT}(s_d,s_d')$; the second equality follows by algebra; the third equality holds because D is fully determined by S (since S being equal to s_d or

†: carol.caetano@uga.edu, ‡: gregorio.caetano@uga.edu, §: brantly.callaway@uga.edu, ¶: ddyal@uga.edu. All authors are affiliated with the John Munro Godfrey, Sr. Department of Economics, University of Georgia.

s'_d implies that D is equal to d); the fourth equality holds by Assumption 2 (or Assumption 6); the fifth equality holds again because S fully determines D; and the sixth equality holds by writing the potential outcomes into their observed counterparts. $\square$

Proof of Lemma C.1. Take any $s_{d-1} \in \mathcal{S}_{d-1}$ such that $d > 0$ where there are K total sub-treatments, i.e. $\mathcal{S}_{d-1} \subset \mathbb{Z}^K_{\geq 0}$ for all $d > 0$. By definition of $\mathcal{S}_{d-1}$, $||s_{d-1}||_1 = d - 1$. Next, add one to any coordinate i of the vector s_d where the i^{th} element in s_{d-1} is not at the maximum of the support of the i^{th} sub-treatment. Denote this vector as $s^* := s_{d-1} + 1_i$, where 1_i is the unit vector of the i^{th} coordinate. Without loss of generality, let $i = 1$. This implies that the ℓ_1 norm of s^* is:

$$||s^*||_1 = \sum_{k=1}^{K} |s^*_k| = \sum_{k=1}^{K} s^*_k = (s_{d-1,1} + 1) + s_{d-1,2} + \cdots + s_{d-1,K} = ||s_{d-1}||_1 + 1 = (d-1) + 1 = d$$

where the first quality holds by the definition of the ℓ_1 norm; the second equality holds since each element of s^* is in the non-negative integers, $\mathbb{Z}^+$; the third equality holds since each element of the vector s^* is the corresponding element of vector s_{d-1}, except the first which is added by one; the fourth equality follows by algebra and the definition of the ℓ_1 norm; the fifth equality holds by the definition of $s_{d-1} \in \mathcal{S}_{d-1}$; and the last equality holds by addition. Hence, $s^* \in \mathcal{S}_d$ by the definition of $\mathcal{S}_d$, and we may write $s_d = s^*$.

Now, recall that we defined without loss of generality that $s^* = s_{d-1} + 1_i = s_d$. This implies that $s_d \succ^+ s_{d-1}$ by the definition of congruent sub-treatment vectors. Therefore, for any $s_{d-1} \in \mathcal{S}_{d-1}$ with $d > 0$, there exists an $s_d \in \mathcal{S}_d$ that is congruent to s_{d-1}. $\square$

Proof of Lemma C.2. Since $||s_d||_1 = d > 0$, $\sum_i^K s_{d,i} = d$ by the definition of a member of S_d. This implies there is at least one coordinate j in s_d with $s_{d,j} > 0$. Next, take any other coordinate l such that $l \neq j$. Since $K \geq 2$, any $l \neq j$ can be chosen. Define another vector $s'_d = s_d - 1_j + 1_l$. Since we know that $s_{d,j} \in \mathbb{Z}^+$ and $s_{d,j} > 0$, subtracting one from the j^{th} coordinate is valid. See that the coordinates:

$$s'_{d,j} = s_{d,j} - 1 \geq 0 \qquad s'_{d,l} = s_{d,l} + 1 \geq 0.$$

Therefore, the ℓ_1 norm of s'_d is

$$||s'_d||_1 = \sum_{i=1}^{K} s'_{d,i} = \sum_{i=1}^{K} s_{d,i} - 1 + 1 = \sum_{i=1}^{K} s_{d,i} = d,$$

where the first equality holds by the definition of the ℓ_1 norm on vectors in the positive orthant; the second equality holds by the definition of s'_d; and the fourth equality holds by the definition of $s_d \in S_d$. This implies that $s'_d \in S_d$ also. Lastly, notice that:

$$s'_d = s_d - 1_j + 1_l \iff s_d = s'_d + 1_j - 1_l,$$

which holds by rearrangement. Therefore, for any $s_d \in S_d$ with $d > 0$ and $K \geq 2$, there always exists a $s'_d \in S_d$ with coordinates j, l such that s_d differs from s'_d by exchanging one unit in coordinate l for one unit in coordinate j, as desired. We say that s_d is a *unit exchange* of s'_d. $\square$

Proof of Lemma C.3. Take any $s_d, s'_d \in \mathcal{S}_d \subset \mathbb{Z}^K_{\geq 0}$ with $d > 0$ such that $||s_d - s'_d||_1 > 2$, where at least for some j^{th} and l^{th} coordinates $s_{d,j} < s'_{d,j}$ and $s_{d,l} > s'_{d,l}$. By Lemma C.2, there exists a vector $s^\star_d \in \mathcal{S}_d$ which is a unit exchange of s_d. All that is left to show is that $||s_d - s'_d||_1 > ||s^\star_d - s'_d||_1$. Since $s^\star_d$ is a unit exchange from s_d, then we may write $s^\star_d = s_d + 1_j - 1_l$, where 1_j and 1_l are unit vectors for some coordinates j and l such that $j \neq l$.

Since we know that any $s_{d,i} \in \mathbb{Z}^+$ for all $i \in \{1, \ldots, K\}$, we have

$$\begin{aligned} s_{d,j} < s'_{d,j} &\iff s_{d,j} < s_{d,j} + 1 \leq s'_{d,j} \\ &\iff s_{d,j} - s'_{d,j} < s_{d,j} + 1 - s'_{d,j} \leq 0 \\ &\iff \left|s_{d,j} - s'_{d,j}\right| > \left|s_{d,j} + 1 - s'_{d,j}\right| \geq 0 \end{aligned}$$

and that

$$\begin{aligned} s_{d,l} > s'_{d,l} &\iff s_{d,l} > s_{d,l} - 1 \geq s'_{d,l} \\ &\iff s_{d,l} + s'_{d,l} > s_{d,l} - 1 + s'_{d,l} \geq 0 \\ &\iff \left|s_{d,l} + s'_{d,l}\right| > \left|s_{d,l} - 1 + s'_{d,l}\right| \geq 0, \end{aligned}$$

where in both results the first inequality is given; the second line is by addition or subtraction of one when all elements are in the set of positive integers; and the last line holds by addition or subtraction of the element $s'_{d,i}$ where $i \in \{j, l\}$. We will use these facts below. Next, see that:

$$\begin{aligned} ||s^\star_d - s'_d||_1 = ||s_d + 1_j - 1_l - s'_d||_1 &= \sum_{i=1}^{K} \left|s_{d,i} + 1_{j,i} - 1_{l,i} - s'_{d,i}\right| \\ &= \sum_{i\neq j\ i\neq l}^{K} \left|(s_{d,i} - s'_{d,i})\right| + \left|(s_{d,j} + 1 - s'_{d,j})\right| + \left|(s_{d,l} - 1 - s'_{d,l})\right| \\ &< \sum_{i\neq j\ i\neq l}^{K} \left|(s_{d,i} - s'_{d,i})\right| + \left|(s_{d,j} - s'_{d,j})\right| + \left|(s_{d,l} - s'_{d,l})\right| = \sum_{i=1}^{K} \left|s_{d,i} - s'_{d,i}\right| = ||s_d - s'_d||_1, \end{aligned}$$

where the first equality holds by the unit exchange property from Lemma C.2; the second equality holds by the definition of the ℓ_1-norm on $\mathbb{Z}^K_{\geq 0}$; the third equality holds by the expansion of the sum; the inequality in the fourth line holds by the facts from above; the fifth equality holds by the rearrangement of the sum; and the sixth equality holds by the definition of the ℓ_1-norm again. This shows that $||s_d - s'_d||_1 > ||s^\star_d - s'_d||_1$. Hence, there always exists a vector $s^\star_d$ which is a unit-exchange away towards the terminating vector s'_d when $||s_d - s'_d||_1 > 2$. $\square$

Lemma SA.1 (Even Distances)**.** *For a fixed $s_d \in \mathcal{S}_d$, take any $s_d^\star \in \mathcal{S}_d$ with $d > 0$ and $K \geq 2$. The ℓ_1-norm distance $0 \leq ||s_d^\star - s_d||_1 = 2 \cdot m \leq 2 \cdot d$, for some $m \in \mathbb{Z}^+$ such that $m \leq d$. That is, any comparison of sub-treatment vectors with the ℓ_1-norm in the same aggregation set always results in an even number between 0 and $2 \cdot d$.*

Proof of Lemma SA.1. For a fixed $s_d \in \mathcal{S}_d$, take any element $s_d^\star \in \mathcal{S}_d$ with $d > 0$ and $K \geq 2$. Since both elements belong to $\mathcal{S}_d$,

$$||s_d^\star||_1 = ||s_d||_1 = d \iff \sum_{k=1}^{K} s_{k,d} = \sum_{k=1}^{K} s_{k,d}^\star = d$$

since all $s_{k,d} \in \mathbb{Z}^+$. However, $s_d \neq s_d^\star$ necessarily. This implies some elements exist such that $s_{d,j}^\star < s_{d,j}$ and $s_{d,l}^\star > s_{d,l}$. Hence:

$$\begin{aligned}
||s_d^\star||_1 - ||s_d||_1 = 0 \iff& \sum_{k=1}^{K} s_{d,k}^\star - \sum_{k=1}^{K} s_{d,k} = 0 \iff \sum_{k=1}^{K} \left(s_{d,k}^\star - s_{d,k}\right) = 0 \\
\iff& \sum_{k:s_{d,k}^\star > s_{d,k}}^{K} \left(s_{d,k}^\star - s_{d,k}\right) + \sum_{k:s_{d,k}^\star < s_{d,k}}^{K} \left(s_{d,k}^\star - s_{d,k}\right) = 0 \\
\iff& \sum_{k:s_{d,k}^\star > s_{d,k}}^{K} \left(s_{d,k}^\star - s_{d,k}\right) = \sum_{k:s_{d,k}^\star < s_{d,k}}^{K} \left(s_{d,k} - s_{d,k}^\star\right) \qquad \text{(S1)}
\end{aligned}$$

where the first line holds by the definition of $\mathcal{S}_d$; the second line holds by the definition of the ℓ_1-norm for numbers belonging to the non-negative integers; the third line holds by rearranging the sum; the fourth line holds since some elements of $s_d^\star$ are not equal to s_d; and the fifth line follows by algebra. This states that the total amount of units in $s_d^\star$ that are greater than the units in s_d must be equal to the number of units in $s_d^\star$ that are less than the number of units in s_d. Also note that both of these terms are positive:

$$\sum_{k:s_{d,k}^\star > s_{d,k}}^{K} \left(s_{d,k}^\star - s_{d,k}\right) > 0 \qquad \sum_{k:s_{d,k}^\star < s_{d,k}}^{K} \left(s_{d,k} - s_{d,k}^\star\right) > 0.$$

Therefore, the ℓ_1-norm:

$$\begin{aligned}
||s_d^\star - s_d||_1 &= \sum_{k=1}^{K} \left|s_{d,k}^\star - s_{d,k}\right| = \sum_{k:s_{d,k}^\star \neq s_{d,k}} \left|s_{d,k}^\star - s_{d,k}\right| \\
&= \sum_{k:s_{d,k}^\star > s_{d,k}} \left(s_{d,k}^\star - s_{d,k}\right) + \sum_{k:s_{d,k}^\star < s_{d,k}} \left(s_{d,k}^\star - s_{d,k}\right) \\
&= 2 \cdot \sum_{k:s_{d,k}^\star > s_{d,k}} \left(s_{d,k}^\star - s_{d,k}\right),
\end{aligned}$$

where the first line is by definition of the ℓ_1-norm; the second line holds since there will be a remainder for all elements not equal between the two vectors; the third line follows by algebra; and the last line holds by Equation (S1) above. We know that $\sum_{k:s^\star_{d,k}>s_{d,k}}\left(s^\star_{d,k}-s_{d,k}\right)\in\mathbb{Z}^+$ as well. Therefore, $||s^\star_d - s_d||_1$ is even always. Moreover, notice the distance at any two $s^\star_d, s_d \in \mathcal{S}_d$ is bound by $2\cdot d$ from the definition of $\mathcal{S}_d$; because the furthest two vectors could be in this measure of distance is where one vector has mass d in a coordinate which is zero in the other vector. Lastly, see that by the Reverse Triangle Inequality,

$$||s^\star_d - s_d||_1 \geq \big|||s^\star_d||_1 - ||s_d||_1\big| = \big|d-d\big| = 0,$$

where $||s^\star_d||_1 = ||s_d||_1 = d$ since they are elements of $\mathcal{S}_d$. Thus, the distance function $g(s^\star_d\big|s_d) := ||s^\star_d - s_d||_1$ is non-negative, bounded by $2\cdot d$, and must result in an even integer. $\square$

Proof of Proposition C.4. Let $s_d, s'_d \in \mathcal{S}_d \subset \mathbb{Z}^K_{\geq 0}$ for $d>0$ and $K\geq 2$. We show that there exists a finite sequence of vectors in $\mathcal{S}_d$ which begins at s_d and terminates at s'_d.

We know by Lemma C.2 that there always exists a unit exchange for any $s_d\in\mathcal{S}_d$. Thus, there exists a vector $s^\star_d \in S_d$ such that $s^\star_d = s_d + 1_j - 1_l$ for some coordinates j and l. By Lemma C.3, there exists a vector $s^\star_d = s^{(1)}_d$ such that $||s_d - s'_d||_1 > ||s^{(1)}_d - s'_d||_1$ unless $s^{(1)}_d = s'_d$. So, there always exists a vector that is a unit exchange from the initial vector that is a shorter distance away from the terminating vector. If $s^{(1)}_d = s'_d$, the claim holds trivially. Next, we construct the process for when $s^{(1)}_d \neq s'_d$.

Initialize $s^{(0)}_d = s_d$. While $||s^{(b)}_d - s'_d||_1 > 0$ for index $b\in\mathbb{Z}$:

1. Pick coordinates j and l such that the corresponding elements $s^{(b)}_{d,j} < s'_{d,j}$ and $s^{(b)}_{d,l} > s'_{d,l}$.
2. Update by the unit exchange property (Lemma C.2): $s^{(b+1)}_d = s^{(b)}_d + 1_j - 1_l$.
3. If $||s^{(b+1)}_d - s'_d||_1 > 0$, store $s^{(b+1)}_d$ and iterate this process.

   Otherwise, if $||s^{(b+1)}_d - s'_d||_1 = 0$, stop the process.

Here, we show that the process terminates and is finite. We know, by Lemmas C.3 and SA.1, that at each step, $0 \leq ||s^{(b+1)}_d - s'_d||_1 < ||s^{(b)}_d - s'_d||_1 \leq 2\cdot d$. Also note that at each step, the distance $||s^{(b+1)}_d - s'_d||_1 - ||s^{(b)}_d - s'_d||_1$ between neighboring vectors relative to the terminating vector s'_d in the chain decreases by two:

$$\begin{aligned}
||s^{(b+1)}_d - s'_d||_1 - ||s^{(b)}_d - s'_d||_1 &= \sum_{i=1}^K \big|s^{(b+1)}_{d,i} - s'_{d,i}\big| - \sum_{i=1}^K \big|s^{(b)}_{d,i} - s'_{d,i}\big| \\
&= \sum_{i\neq j, i\neq l}^K \big|s^{(b)}_{d,i} - s'_{d,i}\big| + \big|s^{(b+1)}_{d,j} - s'_{d,j}\big| - \big|s^{(b+1)}_{d,l} - s'_{d,l}\big| \\
&\qquad - \sum_{i=1}^K \big|s^{(b)}_{d,i} - s'_{d,i}\big|
\end{aligned}$$

$$= \sum_{i\neq j, i\neq l}^{K} \left|s_{d,i}^{(b)} - s'_{d,i}\right| + \left|s_{d,j}^{(b)} + 1 - s'_{d,j}\right| - \left|s_{d,l}^{(b)} - 1 - s'_{d,l}\right|$$

$$- \sum_{i=1}^{K} \left|s_{d,i}^{(b)} - s'_{d,i}\right|$$

$$= \sum_{i\neq j, i\neq l}^{K} \left|s_{d,i}^{(b)} - s'_{d,i}\right| + \left|s_{d,j}^{(b)} - s'_{d,j}\right| + 1 - \left|s_{d,l}^{(b)} - s'_{d,l}\right| + 1$$

$$- \sum_{i=1}^{K} \left|s_{d,i}^{(b)} - s'_{d,i}\right|$$

$$= \sum_{i=1}^{K} \left|s_{d,i}^{(b)} - s'_{d,i}\right| + 2 - \sum_{i=1}^{K} \left|s_{d,i}^{(b)} - s'_{d,i}\right| = 2,$$

where the first equality holds by the definition of the ℓ_1-norm; the second equality holds by rearrangement of the sum; the third equality holds by the definition of $s_d^{(b+1)}$ as a unit exchange of $s_d^{(b)}$; the fourth equality holds by properties of the absolute value; and the fifth and sixth equalities hold by algebra. This demonstrates that each iteration of the process decreases the ℓ_1-norm distance by two.

Since the measure $||s_d^{(b)} - s'_d||_1$ for all possible $s_d^{(b)} \in \mathcal{S}_d$ is non-negative, bound by $2 \cdot d$, always produces an even number (Lemma SA.1) and decreasing monotonically by two at each step, the process will eventually converge to exactly zero. Recall the distance between any two linked vectors in the chain is always two because they are always a unit exchange of one another. Furthermore, recall the largest distance between two vectors in the set $\mathcal{S}_d$ amounts to $2 \cdot d$ for some $d \in \mathbb{Z}^+$, and each distance at each step must be even. Thus, decrements of two from an even number $r = ||s_d - s'_d||_1 = 2 \cdot m$ for some $m \in \mathbb{Z}^+$ such that $m < d$ and $0 < r \leq 2 \cdot d$ will eventually reach zero, which terminates this process. By that, the number of iterations in this process will be equal to $m < d < \infty$, which is finite. This implies that the final vector in the chain is $s_d^{(B)} = s'_d$, as desired, where $B \in \mathbb{Z}^+$ and $B < \infty$. Hence, this process terminates and results in a finite sequence of vectors $\{s_d^{(0)} = s_d, s_d^{(1)}, \ldots, s_d^{(B)} = s'_d\}$. □

Proof of Lemma C.4. Take any $s_{d-1}, s'_{d-1} \in \mathcal{S}_d$ such that $s_{d-1} = s'_{d-1} + 1_j - 1_l$. By definition of congruence, a vector in $\mathcal{S}_d$ is congruent to a vector in S_{d-1} if and only if $s_d = s_{d-1} + 1_k$ for some coordinate k where 1_k is the unit vector in the k^{th} coordinate. Since S_d is not empty, by Lemma C.1, there always exists a congruent vector s_d of $s_{d-1} \in \mathcal{S}_{d-1}$. Notice that

$$s_{d-1} = s'_{d-1} + 1_j - 1_l \iff s_{d-1} + 1_l = s'_{d-1} + 1_j \iff s_d = s'_d,$$

where the first equality holds by assumption; the second equality follows by algebra; and the third equality holds by the definition of a vector in $\mathcal{S}_d$. Note that $s_d \succ^+ s_{d-1}$ (is congruent to s_{d-1}), and $s'_d \succ^+ s'_{d-1}$ (is congruent to s'_{d-1}) by definition of congruency. But, $s_d = s'_d$. Hence, $s_d \succ^+ s_{d-1}$

and $s_d \succ^+ s'_{d-1}$. This implies there is one $s_d \in \mathcal{S}_d$ that is congruent with both s_{d-1} and s'_{d-1}, as desired. Note this vector can be found in two ways: (i) by adding one to the vector s_{d-1} in the corresponding coordinate l; or (ii) by adding one to the vector s'_{d-1} in the corresponding coordinate j. $\square$

Proof of Lemma C.5. We prove Lemma C.5 by induction, for any $K \in \mathbb{Z}^+$ such that $K \geq 1$, that $\sum_{d=1}^{K} \binom{K}{d} \cdot \binom{K}{d-1}$ is equivalent to $\binom{2K}{K-1}$. First, the base case is when $K = 1$. We have that:

$$\sum_{d=1}^{K} \binom{K}{d} \cdot \binom{K}{d-1} = \binom{1}{1} \cdot \binom{1}{0} = 1,$$

and for the claim:

$$\binom{2K}{K-1} = \binom{2}{1-1} = 1$$

as desired. So, the claim holds true for $K = 1$. And, for $K = 2$, the sum is:

$$\sum_{d=1}^{K} \binom{K}{d} \cdot \binom{K}{d-1} = \binom{2}{1} \cdot \binom{2}{0} + \binom{2}{2} \cdot \binom{2}{1} = (2) \cdot (1) + (1) \cdot (2) = 4,$$

and for the claim:

$$\binom{2K}{K-1} = \binom{2\dot{2}}{2-1} = \binom{4}{1} = 4$$

as desired. So, the claim holds true for $K = 2$. Next, we make the induction hypothesis (IH) that, for any $K = k$, the result holds true:

$$\sum_{d=1}^{k} \binom{k}{d} \cdot \binom{k}{d-1} = \binom{2k}{k-1}. \tag{IH}$$

Next, we take the induction step. Take $K = k+1 \in \mathbb{Z}^+$. To show that $\sum_{d=1}^{k+1} \binom{k+1}{d} \cdot \binom{k+1}{d-1} = \binom{2(k+1)}{(k+1)-1}$, see that on the LHS, we have:

$$\begin{aligned}
\sum_{d=1}^{k+1} \binom{k+1}{d} \cdot \binom{k+1}{d-1} &= \sum_{d=1}^{k+1} \left[\binom{k}{d-1} + \binom{k}{d}\right] \cdot \left[\binom{k}{d-2} + \binom{k}{d-1}\right] \\
&= \sum_{d=1}^{k+1} \left[\binom{k}{d-1} \cdot \binom{k}{d-2} + \binom{k}{d-1} \cdot \binom{k}{d-1} + \binom{k}{d} \cdot \binom{k}{d-2} + \binom{k}{d} \cdot \binom{k}{d-1}\right] \\
&= \underbrace{\sum_{d=1}^{k+1} \binom{k}{d-1} \cdot \binom{k}{d-2}}_{\text{(I)}} + \underbrace{\sum_{d=1}^{k+1} \binom{k}{d-1} \cdot \binom{k}{d-1}}_{\text{(II)}}
\end{aligned}$$

$$+\underbrace{\sum_{d=1}^{k+1}\binom{k}{d}\cdot\binom{k}{d-2}}_{\text{(III)}}+\underbrace{\sum_{d=1}^{k+1}\binom{k}{d}\cdot\binom{k}{d-1}}_{\text{(IV)}},$$

where the first equality holds by *binomial recursion*; the second equality comes from algebra; and the third equality holds by the linearity property of the summation operator. We will deal with each part (I)-(IV) separately.

(I) Observe that:

$$\sum_{d=1}^{k+1}\binom{k}{d-1}\cdot\binom{k}{d-2}=\sum_{m=0}^{k}\binom{k}{m}\cdot\binom{k}{m-1}=\sum_{m=1}^{k}\binom{k}{m}\cdot\binom{k}{m-1}=\binom{2k}{k-1},$$

where we shift the index to $m = d - 1$ in the first equality; we drop the $m = 0$ case since that term is zero in the second equality; and in the third equality, we apply the inductive hypothesis (IH).

(II) We have:

$$\begin{aligned}\sum_{d=1}^{k+1}\binom{k}{d-1}\cdot\binom{k}{d-1}&=\sum_{m=0}^{k}\binom{k}{d-1}\cdot\binom{k}{d-1}\\&=\sum_{m=0}^{k}\binom{k}{m}\cdot\frac{k!}{m!(k-m)!}\\&=\sum_{m=0}^{k}\binom{k}{m}\cdot\frac{k!}{(k-m)!(k-(k-m))!}\\&=\sum_{m=0}^{k}\binom{k}{m}\cdot\binom{k}{k-m}=\binom{2k}{k},\end{aligned}$$

where we shift the index to $m = d-1$ in the first equality; the second equality is by definition of a binomial coefficient; the third equality is by algebra; and in the fourth equality, we apply *Vandermonde's identity* from combinatorics.

(III) Note that:

$$\begin{aligned}\sum_{d=1}^{k+1}\binom{k}{d}\cdot\binom{k}{d-2}&=\sum_{d=1}^{k}\binom{k}{d}\cdot\binom{k}{d-2}=\sum_{i=0}^{k-2}\binom{k}{i+2}\cdot\binom{k}{i}=\sum_{i=0}^{k-2}\binom{k}{i}\cdot\frac{k!}{(i+2)!(k-(i+2))!}\\&=\sum_{i=0}^{k-2}\binom{k}{i}\cdot\frac{k!}{(k-(i+2))!(k-(k-(i+2))!}=\sum_{i=0}^{k-2}\binom{k}{i}\cdot\binom{k}{k-(i+2)}\\&=\sum_{i=0}^{k-2}\binom{k}{i}\cdot\binom{k}{(k-2)-i}=\binom{k+k}{k-2}=\binom{2k}{k-2},\end{aligned}$$

where the first equality drops the index from $k + 1$ to k since that term will be zero; we shift the index to $i = d-2$ in the second equality since the first two terms will be zero in the sum; the third, fourth, fifth, and sixth equalities are by algebra; and the seventh equality is by *Vandermonde's identity*.

(IV) Note that:

$$\sum_{d=1}^{k+1} \binom{k}{d} \cdot \binom{k}{d-1} = \sum_{d=1}^{k} \binom{k}{d} \cdot \binom{k}{d-1} = \binom{2k}{k-1},$$

where the first equality happens since at $k+1$ the term is zero; and the second equality arises by the induction hypothesis (IH).

Now, by (I)-(IV), we have:

$$\begin{aligned}
\sum_{d=1}^{k+1} \binom{k+1}{d} \cdot \binom{k+1}{d-1} &= \binom{2k}{k-1} + \binom{2k}{k} + \binom{2k}{k-2} + \binom{2k}{k-1} \\
&= \left[\binom{2k}{k-1} + \binom{2k}{k}\right] + \left[\binom{2k}{k-2} + \binom{2k}{k-1}\right] \\
&= \binom{2k+1}{k} + \binom{2k+1}{k-1} \\
&= \binom{2k+2}{k} = \binom{2(k+1)}{(k+1)-1},
\end{aligned}$$

as desired. Note that the third and fourth equalities are by the binomial recursive (Pascal's) identity; and the fifth equality is by algebra. Thus, we have shown that the claim holds for any K. $\square$

Proof of Proposition C.6. Given $K > 1$ trinary sub-treatments, we show that for all $d \in \mathcal{D}$: (i) the total number of distinct disaggregated contrasts is $\sum_{d=1}^{K} \sum_{r=0}^{\lfloor d/2 \rfloor} \binom{K}{r} \cdot \binom{K-r}{d-2r}$; and (ii) the number of congruent disaggregated contrasts is the row-sum of a row-scaled extended trinomial triangle, whose elements are the sum of the top-three adjacent elements.

1. *Total distinct contrasts:* First, we show that the number of K-tuples that sum to $D = d$ is $\sum_{r=0}^{\lfloor d/2 \rfloor} \binom{K}{r} \cdot \binom{K-r}{d-2r}$. The size of a tuple is K. Since each variable is trinary, each element in the K-tuple is either zero, one or two. This turns out to be equivalent to the elements of the trinomial triangle, where the rows are K and the columns are d. The trinomial triangle is similar to Pascal's triangle except that each entry is the sum of the above three adjacent entries. The trinomial triangle takes the following form:

$K=0$:					1					
$K=1$:				1	1	1				
$K=2$:			1	2	3	2	1			
$K=3$:		1	3	6	7	6	3	1		
$K=4$:	1	4	10	16	19	16	10	4	1	

where it is known that each $(K,d)^{th}$ entry is equal to the trinomial coefficient $\binom{K}{d}_2 := \sum_{r=0}^{\lfloor d/2 \rfloor} \binom{K}{r} \cdot \binom{K-r}{d-2r}$. Hence, if we wish to find the total number of distinct contrasts, we sum over all $d \in \{1, \cdots, K\}$ for the corresponding row (a fixed K), which can be expressed as $\sum_{d}^{K} \binom{K}{d}_2 = \sum_{d=1}^{K} \sum_{r=0}^{\lfloor d/2 \rfloor} \binom{K}{r} \cdot \binom{K-r}{d-2r}$, as desired.

2. *Congruent distinct contrasts:* Next, we show that the number of congruent contrasts for K trinary sub-treatments is the sum of the rows of an extended trinomial triangle multiplied by K, where the rows are indexed by K and the columns are indexed by aggregate $D = d$. Analogously, the elements that make up this triangle are the sum of the three adjacent elements in the row above. The extended trinomial triangle has the following form:

$K=1$:					1	1				
$K=2$:				1	2	2	1			
$K=3$:			1	3	5	5	3	1		
$K=4$:		1	4	9	13	13	9	4	1	
$K=5$:	1	5	14	26	35	35	26	14	5	1

For each K, the number of congruent contrasts can be listed and counted for verification by program.

□

SA.2 Proofs of Results from Section 4

Lemma SA.2 (Moment Equations for Regression of Y on Aggregated D)**.** *For a discrete random variable D with non-negative support $\mathcal{D} := \{0, \ldots, \bar{N}\}$ for some $\bar{N} \in \mathbb{Z}^+$, and any non-degenerate random variable Y, we have the following results:*

$$\mathbb{E}[D] = \sum_{d=0}^{\bar{N}} d \cdot p_d \qquad \mathbb{E}[D^2] = \sum_{d=0}^{\bar{N}} d^2 \cdot p_d \qquad \mathbb{E}[DY] = \sum_{d=0}^{\bar{N}} d \cdot \mathbb{E}_d \cdot p_d \qquad \mathbb{E}[Y] = p_0 \cdot \mathbb{E}_0 + \sum_{d=1}^{\bar{N}} p_d \cdot (\mathbb{E}_d - \mathbb{E}_0),$$

where $p_d := P(D = d)$; and $\mathbb{E}_d := \mathbb{E}[Y | D = d], \forall d \in \mathcal{D}$.

Proof of Lemma SA.2. For the first two equations, we have:

$$\mathbb{E}[D] = \sum_{d \in \mathcal{D}} d \cdot P(D = d) = \sum_{d=0}^{\bar{N}} d \cdot p_d,$$
$$\mathbb{E}[D^2] = \sum_{d \in \mathcal{D}} d^2 \cdot P(D = d) = \sum_{d=0}^{\bar{N}} d^2 \cdot p_d,$$

which are both by the definition of expectation, and where $p_d := P(D = d)$. For the last two equations, we have:

$$\mathbb{E}[DY] = \mathbb{E}[\mathbb{E}[DY|D]] = \sum_{d=0}^{\bar{N}} d \cdot \mathbb{E}[Y|D = d] \cdot P(D = d) = \sum_{d=0}^{\bar{N}} d \cdot \mathbb{E}_d \cdot p_d,$$
$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|D]] = \sum_{d=0}^{\bar{N}} \mathbb{E}[Y|D = d] \cdot P(D = d) = \sum_{d=0}^{\bar{N}} \mathbb{E}_d \cdot p_d = p_0 \cdot \mathbb{E}_0 + \sum_{d=1}^{\bar{N}} p_d \cdot (\mathbb{E}_d - \mathbb{E}_0),$$

which both hold by the law of iterated expectations, and $\mathbb{E}_d := \mathbb{E}[Y|D = d], \forall d \in \mathcal{D}$. □

Lemma SA.3 (Covariance of Outcomes Y and Aggregated Treatment D for Baseline-to-d Primitives)**.** *The covariance between a non-degenerate random variable Y and a discrete random variable D with support $\mathcal{D} = \{0, \ldots, \bar{N}\}$ for some $\bar{N} \in \mathbb{Z}^+$ can be expressed as:*

$$Cov(Y, D) = \mathbb{E}[YD] - \mathbb{E}[Y] \cdot \mathbb{E}[D] = \sum_{d=1}^{\bar{N}} d \cdot w_d \cdot \left(\frac{\mathbb{E}[Y|D = d] - \mathbb{E}[Y|D = 0]}{d} \right),$$

where the weight $w_d = p_d \cdot (d - \mathbb{E}[D])$ represents the weighted distance from the mean of aggregated treatment D.

Proof of Lemma SA.3. With the results from Lemma SA.2, we show the covariance between Y and aggregated treatment D can be expressed as:

$$\begin{aligned}
Cov(Y, D) &= \mathbb{E}[YD] - \mathbb{E}[Y] \cdot \mathbb{E}[D] \\
&= \sum_{d=0}^{\bar{N}} d \cdot \mathbb{E}_d \cdot p_d - \left(\mathbb{E}_0 \cdot p_0 + \sum_{d=1}^{\bar{N}} \mathbb{E}_d \cdot p_d \right) \cdot \mathbb{E}[D] \\
&= \sum_{d=1}^{\bar{N}} d \cdot \mathbb{E}_d \cdot p_d - \sum_{d=1}^{\bar{N}} \mathbb{E}_d \cdot p_d \cdot \mathbb{E}[D] - \mathbb{E}_0 \cdot p_0 \cdot \mathbb{E}[D] \\
&= \sum_{d=1}^{\bar{N}} \mathbb{E}_d \cdot p_d \cdot (d - \mathbb{E}[D]) - \mathbb{E}_0 \cdot p_0 \cdot \mathbb{E}[D] \\
&= \sum_{d=1}^{\bar{N}} \mathbb{E}_d \cdot p_d \cdot (d - \mathbb{E}[D]) - \mathbb{E}_0 \cdot \left(1 - \sum_{d=1}^{\bar{N}} p_d \right) \cdot \mathbb{E}[D]
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{d=1}^{\bar{N}} p_d \cdot (d - \mathbb{E}[D]) \cdot \mathbb{E}_d - \left(\mathbb{E}_0 \cdot \mathbb{E}[D] - \sum_{d=1}^{\bar{N}} p_d \cdot \mathbb{E}_0 \cdot \mathbb{E}[D] \right) \\
&= \sum_{d=1}^{\bar{N}} p_d \cdot ((d - \mathbb{E}[D]) \cdot \mathbb{E}_d + \mathbb{E}_0 \cdot \mathbb{E}[D]) - \mathbb{E}_0 \cdot \mathbb{E}[D] \\
&= \sum_{d=1}^{\bar{N}} p_d \cdot (d \cdot \mathbb{E}_d - \mathbb{E}[D] \cdot (\mathbb{E}_d - \mathbb{E}_0)) - \mathbb{E}_0 \cdot \sum_{d=0}^{\bar{N}} d \cdot p_d \\
&= \sum_{d=1}^{\bar{N}} p_d \cdot (d \cdot \mathbb{E}_d - \mathbb{E}[D] \cdot (\mathbb{E}_d - \mathbb{E}_0) - d \cdot \mathbb{E}_0) \\
&= \sum_{d=1}^{\bar{N}} p_d \cdot (d \cdot (\mathbb{E}_d - \mathbb{E}_0) - \mathbb{E}[D] \cdot (\mathbb{E}_d - \mathbb{E}_0)) \\
&= \sum_{d=1}^{\bar{N}} p_d \cdot (d - \mathbb{E}[D]) \cdot (\mathbb{E}_d - \mathbb{E}_0) := \sum_{d=1}^{\bar{N}} w_d \cdot (\mathbb{E}[Y|D=d] - \mathbb{E}[Y|D=0]) \\
&= \sum_{d=1}^{\bar{N}} d \cdot w_d \cdot \left(\frac{\mathbb{E}[Y|D=d] - \mathbb{E}[Y|D=0]}{d} \right),
\end{aligned}
$$

where the weight $w_d = p_d \cdot (d - \mathbb{E}[D])$, as desired. □

Lemma SA.4 (Overall Weights with Scaled Baseline-to-d Primitives Sum to One)**.** *The weights in Proposition SA.1 from a regression involving scaled baseline-to-d building blocks sum to one.*

Proof of Lemma SA.4. The proof subsumes the proof that the regression weights in Proposition SA.1 involving scaled baseline-to-d building blocks also sum to one. Note that the sum of the overall weights:

$$
\begin{aligned}
\sum_{d=1}^{\bar{N}} \sum_{s \in \mathcal{S}_d} \tilde{\omega}^{reg}(d) \cdot \mathrm{P}(S = s_d | D = d) &= \sum_{d=1}^{\bar{N}} \sum_{s \in \mathcal{S}_d} \frac{d \cdot (d - \mathbb{E}[D])}{Var(D)} \cdot P(D=d) \cdot P(S = s_d | D = d) \\
&= \sum_{d=1}^{\bar{N}} \frac{d \cdot (d - \mathbb{E}[D])}{Var(D)} \cdot P(D=d) \cdot \sum_{s \in \mathcal{S}_d} P(S = s_d | D = d) \\
&= \sum_{d=1}^{\bar{N}} \frac{d \cdot (d - \mathbb{E}[D])}{Var(D)} \cdot P(D=d) \cdot 1 \\
&= \frac{\sum_{d=1}^{\bar{N}} d^2 - d \cdot \mathbb{E}[D]}{Var(D)} \cdot P(D=d) \\
&= \frac{\sum_{d=1}^{\bar{N}} d^2 \cdot P(D=d) - \mathbb{E}[D] \sum_{d=1}^{\bar{N}} d \cdot P(D=d)}{Var(D)} \\
&= \frac{\mathbb{E}[D^2] - \mathbb{E}[D]^2}{Var(D)} = \frac{Var(D)}{Var(D)} = 1,
\end{aligned}
$$

where the first equality holds by definition from Proposition SA.1; the second equality uses the linearity property of the summation operator; the third equality applies the unity axiom of probability; the fourth equality and fifth equality hold by algebra; the sixth equality holds by the definition of expectation; and the seventh equality holds by the definition of variance. $\square$

Lemma SA.5 (Regression Weights with Scaled Baseline-to-d Primitives Are Negative, Positive and Zero)**.** *The regression weights in Proposition SA.1 involving scaled baseline-to-d building blocks have negative, positive, and possibly zero sign. That is,* $sgn(\tilde{\omega}^{reg}(d)) \in \{-1, 0, 1\}$.

Proof of Lemma SA.5. We present here the proof that the regression weights on the scaled-adjusted contrasts derived from Proposition SA.1 can be negative, positive, or zero. See that for any $d \in \mathcal{D}_{>0}$, the regression weight is defined as:

$$\tilde{\omega}^{reg}(d) = \frac{d \cdot (d - \mathbb{E}[D])}{Var(D)} \cdot P(D = d).$$

Notice that, for non-degenerate D, $Var(D) > 0$ and $P(D = d) \geq 0$. Hence, the regression weight can be negative, positive, or zero if:

$$\tilde{\omega}^{reg}(d) \lesseqgtr 0 \iff (d - \mathbb{E}[D]) \lesseqgtr 0 \iff d \lesseqgtr \mathbb{E}[D].$$

That is, the regression weight from using scaled baseline-to-d building blocks is negative if $d < \mathbb{E}[D]$, positive if $d > \mathbb{E}[D]$, and zero if $d = \mathbb{E}[D]$. $\square$

Proposition SA.1. *Whether sub-treatments are observed or unobserved,* α_1 *from the regression in Equation (1) can be decomposed as follows*

$$\alpha_1 = \sum_{d=1}^{\bar{N}} \tilde{\omega}^{reg}(d) \cdot \frac{\mathbb{E}[Y|D = d] - \mathbb{E}[Y|D = 0]}{d},$$

or, written in terms of sub-treatment vectors, as

$$\alpha_1 = \sum_{d=1}^{\bar{N}} \sum_{s_d \in \mathcal{S}_d} \tilde{\omega}^{reg}(d) \cdot \mathrm{P}(S = s_d | D = d) \cdot \left(\frac{\mathbb{E}[Y|S = s_d] - \mathbb{E}[Y|S = 0_K]}{d} \right),$$

where $\tilde{\omega}^{reg}(d) = \frac{d \cdot (d - \mathbb{E}[D])}{\mathrm{Var}(D)} \cdot \mathrm{P}(D = d)$*, which satisfies the following properties:*

(i) $\sum_{d=1}^{\bar{N}} \tilde{\omega}^{reg}(d) = 1$*, (ii)* $\sum_{d=1}^{\bar{N}} \sum_{s_d \in \mathcal{S}_d} \tilde{\omega}^{reg}(d) \cdot \mathrm{P}(S = s_d | D = d) = 1$*, and (iii)* $\tilde{\omega}^{reg}(d) \lesseqgtr 0$ *for* $d \lesseqgtr \mathbb{E}[D]$.

Proof of Proposition SA.1. We show that for any number of multivalued sub-treatments $K \in$

$\mathbb{Z}^+$ using model (1), that the causal parameter α_1 is:

$$\alpha_1 = \sum_{d=1}^{\bar{N}} \tilde{\omega}^{reg}(d) \cdot \frac{\mathbb{E}[Y|D=d] - \mathbb{E}[Y|D=0]}{d}$$
$$= \sum_{d=1}^{\bar{N}} \sum_{s \in \mathcal{S}_d} \tilde{\omega}^{reg}(d) \cdot \left(\mathrm{P}(S = s_d | D = d) \cdot \frac{\mathbb{E}[Y|S = s_d] - \mathbb{E}[Y|S = 0_K]}{d} \right).$$

Choose any $K \in \mathbb{Z}^+$ to be the number of discrete treatments with countably positive finite support, $\mathcal{S}_1 := \{0, 1, \ldots, N_1\}$, $\mathcal{S}_2 = \{0, 1, \ldots, N_2\}, \ldots, \mathcal{D}_K = \{0, 1, \ldots, N_K\}$. Denote the largest possible level of aggregated treatment as $\bar{N} := \sum_{k=1}^{K} N_k$. Our aggregated treatment variable is $D := \sum_{k=1}^{K} S_k$ with support $\mathcal{D} := \{0, 1, \ldots, \bar{N}\}$. Assuming model (1), the causal parameter α_1 is:

$$\alpha_1 = \frac{Cov(Y, D)}{Var(D)} = \frac{\mathbb{E}[YD] - \mathbb{E}[Y]\mathbb{E}[D]}{\mathbb{E}[D^2] - \mathbb{E}[D]^2}.$$

By application of Lemmas SA.2 and SA.3 into the numerator of α_1, we obtain the aggregate parameter:

$$\alpha_1 = \frac{\sum_{d=1}^{\bar{N}} d \cdot w_d \cdot \left(\frac{\mathbb{E}[Y|D=d] - \mathbb{E}[Y|D=0]}{d} \right)}{Var(D)}$$
$$= \frac{\sum_{d=1}^{\bar{N}} d \cdot (p_d \cdot (d - \mathbb{E}[D])) \cdot \left(\frac{\mathbb{E}[Y|D=d] - \mathbb{E}[Y|D=0]}{d} \right)}{Var(D)}$$
$$= \sum_{d=1}^{\bar{N}} \frac{d \cdot (d - \mathbb{E}[D])}{Var(D)} \cdot P(D = d) \cdot \left(\frac{\mathbb{E}[Y|D = d] - \mathbb{E}[Y|D = 0]}{d} \right)$$
$$:= \sum_{d=1}^{\bar{N}} \tilde{\omega}^{reg}(d) \cdot \frac{\mathbb{E}[Y|D = d] - \mathbb{E}[Y|D = 0]}{d}, \tag{S2}$$

where the weights inherited from the regression are $\tilde{\omega}^{reg}(d) = \frac{d \cdot (d - \mathbb{E}[D])}{Var(D)} \cdot P(D = d)$, as we had sought. Notice, by SA.5, that the regression weight will be negative for $d < \mathbb{E}[D]$, positive for $d > \mathbb{E}[D]$, and zero if $d = \mathbb{E}[D]$.

Next, we expand the result in (S2) to the level of sub-treatments. Denote the vector of sub-treatment variables as $s \in \mathcal{S} := \bigtimes_{k=1}^{K} \mathcal{S}_k$, where $\mathcal{S}$ is the Cartesian product of all sub-treatment supports. Denote the vector of realized sub-treatment values that have elements summing to $D = d$ as s_d. It is sufficient to show that for all $s \in \mathcal{S}_d$ and all $d \in \mathcal{D}$, the difference $\mathbb{E}[Y|D = d] - \mathbb{E}[Y|D = 0]$ can be expressed as:

$$\mathbb{E}[Y|D = d] - \mathbb{E}[Y|D = 0] = \sum_{s \in \mathcal{S}_d} P(S = s_d | D = d) \cdot \mathbb{E}[Y|S = s_d] - \mathbb{E}[Y|D = 0]$$
$$= \sum_{s \in \mathcal{S}_d} P(S = s_d | D = d) \cdot \mathbb{E}[Y|S = s_d] - \left(\sum_{s \in \mathcal{S}_d} P(S = s_d | D = d) \right) \cdot \mathbb{E}[Y|D = 0]$$

$$= \sum_{s \in \mathcal{S}_d} P(S = s_d | D = d) \cdot (\mathbb{E}[Y|S = s_d] - \mathbb{E}[Y|D = 0]) \,. \tag{S3}$$

Plugging expression (S3) into (S2) produces:

$$\begin{aligned}
\alpha_1 &= \sum_{d=1}^{\bar{N}} \tilde{\omega}^{reg}(d) \cdot \frac{\mathbb{E}[Y|D = d] - \mathbb{E}[Y|D = 0]}{d} \\
&= \sum_{d=1}^{\bar{N}} \tilde{\omega}^{reg}(d) \cdot \frac{\sum_{s \in \mathcal{S}_d} P(S = s_d|D = d) \cdot (\mathbb{E}[Y|S = s_d] - \mathbb{E}[Y|D = 0])}{d} \\
&= \sum_{d=1}^{\bar{N}} \sum_{s \in \mathcal{S}_d} \tilde{\omega}^{reg}(d) \cdot P(S = s_d|D = d) \cdot \left(\frac{\mathbb{E}[Y|S = s_d] - \mathbb{E}[Y|D = 0]}{d} \right),
\end{aligned}$$

as desired, where the weights satisfy the following properties: (i) $\sum_{d=1}^{\bar{N}} \tilde{\omega}^{reg}(d) = 1$ (Lemma SA.4); (ii) $\sum_{d=1}^{\bar{N}} \sum_{s_d \in \mathcal{S}_d} \tilde{\omega}^{reg}(d) \cdot \mathrm{P}(S = s_d|D = d) = 1$ (Lemma SA.4); and $\tilde{\omega}^{reg}(d) \lessgtr 0$ for $d \lessgtr \mathbb{E}[D]$ (Lemma SA.5). □

SA.3 Proofs of Results from Appendix B.2

Lemma SA.6 (Covariance of Outcome Y with Aggregated D)**.** *Following the results from Lemma SA.2, we have that the covariance between any non-degenerate random variable Y and a non-negative, discrete random variable D with support $\mathcal{D} := \{0, \ldots, \bar{N}\}$ for some $\bar{N} \in \mathbb{Z}^+$, can be written as:*

$$Cov(Y, D) = \mathbb{E}[YD] - \mathbb{E}[Y] \cdot \mathbb{E}[D] = \sum_{d=1}^{\bar{N}} \left(\sum_{j=d}^{\bar{N}} w_j \right) \cdot \Big(\mathbb{E}[Y|D = d] - \mathbb{E}[Y|D = d - 1] \Big),$$

where $w_j := p_j \cdot (j - \mathbb{E}[D])$, for all $j \in \mathcal{D}$.

Proof of Lemma SA.6. Following the results from Lemma SA.2, see that the covariance between the outcome Y and aggregated treatment D can be written as:

$$\begin{aligned}
Cov(Y, D) &= \mathbb{E}[YD] - \mathbb{E}[Y] \cdot \mathbb{E}[D] \\
&= \sum_{d=0}^{\bar{N}} d \cdot \mathbb{E}_d \cdot p_d - \left(\mathbb{E}_0 \cdot p_0 + \sum_{d=1}^{\bar{N}} \mathbb{E}_d \cdot p_d \right) \cdot \mathbb{E}[D] \\
&= \sum_{d=1}^{\bar{N}} d \cdot \mathbb{E}_d \cdot p_d - \sum_{d=1}^{\bar{N}} \mathbb{E}_d \cdot p_d \cdot \mathbb{E}[D] - \mathbb{E}_0 \cdot p_0 \cdot \mathbb{E}[D] \\
&= \sum_{d=1}^{\bar{N}} \mathbb{E}_d \cdot p_d \cdot (d - \mathbb{E}[D]) - \mathbb{E}_0 \cdot p_0 \cdot \mathbb{E}[D]
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{d=1}^{\bar{N}} \mathbb{E}_d \cdot p_d \cdot (d - \mathbb{E}[D]) - \mathbb{E}_0 \cdot \left(1 - \sum_{d=1}^{\bar{N}} p_d\right) \cdot \mathbb{E}[D] \\
&= \sum_{d=1}^{\bar{N}} p_d \cdot (d - \mathbb{E}[D]) \cdot \mathbb{E}_d - \left(\mathbb{E}_0 \cdot \mathbb{E}[D] - \sum_{d=1}^{\bar{N}} p_d \cdot \mathbb{E}_0 \cdot \mathbb{E}[D]\right) \\
&= \sum_{d=1}^{\bar{N}} p_d \cdot ((d - \mathbb{E}[D]) \cdot \mathbb{E}_d + \mathbb{E}_0 \cdot \mathbb{E}[D]) - \mathbb{E}_0 \cdot \mathbb{E}[D] \\
&= \sum_{d=1}^{\bar{N}} p_d \cdot (d \cdot \mathbb{E}_d - \mathbb{E}[D] \cdot (\mathbb{E}_d - \mathbb{E}_0)) - \mathbb{E}_0 \cdot \sum_{d=0}^{\bar{N}} d \cdot p_d \\
&= \sum_{d=1}^{\bar{N}} p_d \cdot (d \cdot \mathbb{E}_d - \mathbb{E}[D] \cdot (\mathbb{E}_d - \mathbb{E}_0) - d \cdot \mathbb{E}_0) \\
&= \sum_{d=1}^{\bar{N}} p_d \cdot (d \cdot (\mathbb{E}_d - \mathbb{E}_0) - \mathbb{E}[D] \cdot (\mathbb{E}_d - \mathbb{E}_0)) \\
&= \sum_{d=1}^{\bar{N}} p_d \cdot (d - \mathbb{E}[D]) \cdot (\mathbb{E}_d - \mathbb{E}_0) := \sum_{d=1}^{\bar{N}} w_d \cdot (\mathbb{E}_d - \mathbb{E}_0),
\end{aligned}
\tag{S4}
$$

where $p_d := P(D = d)$; and $\mathbb{E}_d := \mathbb{E}[Y|D = d], \forall d \in \mathcal{D}$. From (S4), we can further decompose the sum into incremental increases in aggregated treatment intensity:

$$
\begin{aligned}
\sum_{d=1}^{\bar{N}} w_d \cdot (\mathbb{E}_d - \mathbb{E}_0) &= w_1 \cdot (\mathbb{E}_1 - \mathbb{E}_0) + \sum_{d=2}^{\bar{N}} w_d \cdot (\mathbb{E}_d - \mathbb{E}_0 + \mathbb{E}_{d-1} - \mathbb{E}_{d-1}) \\
&= w_1 \cdot (\mathbb{E}_1 - \mathbb{E}_0) + w_1 \cdot (\mathbb{E}_2 - \mathbb{E}_0) + \sum_{d=2}^{\bar{N}} w_d \cdot (\mathbb{E}_d - \mathbb{E}_{d-1}) + \sum_{d=2}^{\bar{N}} w_d \cdot (\mathbb{E}_{d-1} - \mathbb{E}_0) \\
&= (w_1 + w_2) \cdot (\mathbb{E}_1 - \mathbb{E}_0) + \sum_{d=2}^{\bar{N}} w_d \cdot (\mathbb{E}_d - \mathbb{E}_{d-1}) + \sum_{d=3}^{\bar{N}} w_d \cdot (\mathbb{E}_{d-1} - \mathbb{E}_0 + \mathbb{E}_{d-2} - \mathbb{E}_{d-2}) \\
&= (w_1 + w_2) \cdot (\mathbb{E}_1 - \mathbb{E}_0) + \sum_{d=2}^{\bar{N}} w_d \cdot (\mathbb{E}_d - \mathbb{E}_{d-1}) \\
&\quad + \sum_{d=3}^{\bar{N}} w_d \cdot (\mathbb{E}_{d-1} - \mathbb{E}_{d-2}) + \sum_{d=3}^{\bar{N}} w_d \cdot (\mathbb{E}_{d-2} - \mathbb{E}_0) \\
&= (w_1 + w_2 + \cdots) \cdot (\mathbb{E}_1 - \mathbb{E}_0) + \sum_{d=2}^{\bar{N}} w_d \cdot (\mathbb{E}_d - \mathbb{E}_{d-1}) + \sum_{d=3}^{\bar{N}} w_d \cdot (\mathbb{E}_{d-1} - \mathbb{E}_{d-2}) \\
&\quad + \cdots + \sum_{d=\bar{N}-1}^{\bar{N}} w_d \cdot (\mathbb{E}_{d-(\bar{N}-3)} - \mathbb{E}_{d-(\bar{N}-2)}) + w_{\bar{N}} \cdot (\mathbb{E}_{\bar{N}} - \mathbb{E}_{\bar{N}-1}) \\
&= \sum_{d=1}^{\bar{N}} w_d \cdot (\mathbb{E}_1 - \mathbb{E}_0) + \sum_{d=2}^{\bar{N}} w_d \cdot (\mathbb{E}_d - \mathbb{E}_{d-1}) + \cdots
\end{aligned}
$$

$$+\sum_{d=\bar{N}-1}^{\bar{N}} w_d \cdot (\mathbb{E}_{d-(\bar{N}-3)} - \mathbb{E}_{d-(\bar{N}-2)}) + w_{\bar{N}} \cdot (\mathbb{E}_{\bar{N}} - \mathbb{E}_{\bar{N}-1}). \tag{S5}$$

Notice that, collecting the terms for the difference $(\mathbb{E}_2 - \mathbb{E}_1)$ in (S5), we have:

$$w_2 \cdot (\mathbb{E}_2 - \mathbb{E}_1) + w_3 \cdot (\mathbb{E}_2 - \mathbb{E}_1) + \cdots + w_{\bar{N}} \cdot (\mathbb{E}_2 - \mathbb{E}_1) = \sum_{d=2}^{\bar{N}} w_d \cdot (\mathbb{E}_2 - \mathbb{E}_1).$$

Similarly, for the difference $(\mathbb{E}_3 - \mathbb{E}_2)$ in (S5) we have:

$$w_3 \cdot (\mathbb{E}_3 - \mathbb{E}_2) + w_4 \cdot (\mathbb{E}_3 - \mathbb{E}_2) + \cdots + w_{\bar{N}} \cdot (\mathbb{E}_3 - \mathbb{E}_2) = \sum_{d=3}^{\bar{N}} w_d \cdot (\mathbb{E}_3 - \mathbb{E}_2),$$

and so forth. Hence we may rewrite the expression in (S5) as:

$$\begin{aligned}
&\sum_{d=1}^{\bar{N}} w_d \cdot (\mathbb{E}_1 - \mathbb{E}_0) + \sum_{d=2}^{\bar{N}} w_d \cdot (\mathbb{E}_d - \mathbb{E}_{d-1}) + \cdots + \sum_{d=K-1}^{\bar{N}} w_d \cdot (\mathbb{E}_{d-(\bar{N}-3)} - \mathbb{E}_{d-(\bar{N}-2)}) + w_{\bar{N}} \cdot (\mathbb{E}_{\bar{N}} - \mathbb{E}_{\bar{N}-1}) \\
&= \sum_{d=1}^{\bar{N}} w_d \cdot (\mathbb{E}_1 - \mathbb{E}_0) + \sum_{d=2}^{\bar{N}} w_d \cdot (\mathbb{E}_2 - \mathbb{E}_1) + \sum_{d=3}^{\bar{N}} w_d \cdot (\mathbb{E}_3 - \mathbb{E}_2) + \cdots + w_{\bar{N}} \cdot (\mathbb{E}_{\bar{N}} - \mathbb{E}_{\bar{N}-1}) \\
&= \sum_{d=1}^{\bar{N}} (\mathbb{E}_d - \mathbb{E}_{d-1}) \cdot \sum_{j=d}^{\bar{N}} w_j = \sum_{d=1}^{\bar{N}} \left(\sum_{j=d}^{\bar{N}} w_j \right) \cdot (\mathbb{E}_d - \mathbb{E}_{d-1}) \\
&= \sum_{d=1}^{\bar{N}} \left(\sum_{j=d}^{\bar{N}} w_j \right) \cdot \Big(\mathbb{E}[Y|D=d] - \mathbb{E}[Y|D=d-1] \Big),
\end{aligned}$$

as desired. □

Lemma SA.7 (Variance of Aggregated D)**.** *For a discrete random variable D with non-negative support $\mathcal{D} := \{0, \ldots, \bar{N}\}$ for some $\bar{N} \in \mathbb{Z}^+$, we may express the variance of D as:*

$$Var(D) = \mathbb{E}[D^2] - \mathbb{E}[D]^2 = \sum_{d=1}^{\bar{N}} d \cdot w_d,$$

where $w_d := p_d \cdot (d - \mathbb{E}[D])$.

Proof of Lemma SA.7.

$$\begin{aligned}
Var(D) &= \mathbb{E}[D^2] - \mathbb{E}[D]^2 \\
&= \sum_{d=0}^{\bar{N}} d^2 \cdot p_d - \left(\sum_{d=0}^{\bar{N}} d \cdot p_d \right) \cdot \mathbb{E}[D] = \sum_{d=1}^{\bar{N}} d^2 \cdot p_d - \left(\sum_{d=1}^{\bar{N}} d \cdot p_d \right) \cdot \mathbb{E}[D] \\
&= \sum_{d=1}^{\bar{N}} (d^2 \cdot p_d - d \cdot p_d \cdot \mathbb{E}[D]) = \sum_{d=1}^{\bar{N}} d \cdot p_d \cdot (d - \mathbb{E}[D]) = \sum_{d=1}^{\bar{N}} d \cdot w_d,
\end{aligned}$$

where the first equality holds by definition; the second equality holds by Lemma SA.2 and by applying the unity property of probability; the third equality holds since the term for $D = 0$ is zero; and the fourth and fifth equality hold by the summation operator and by algebra. □

Lemma SA.8 (Re-expressing the Regression Weight with Marginal Primitives)**.** *The regression weight* $\omega^{reg}(d) = \frac{P(D\geq d)\cdot(\mathbb{E}[D|D\geq d]-\mathbb{E}[D])}{Var(D)}$ *from Proposition B.1 can be written as* $\frac{\sum_{j=d}^{\bar{N}} w_j}{\left(\sum_{d=1}^{\bar{N}} d\cdot w_d\right)}$.

Proof of Lemma SA.8. We show that the quantity $\frac{\sum_{j=d}^{\bar{N}} w_j}{\left(\sum_{d=1}^{\bar{N}} d\cdot w_d\right)}$ can be expressed as the regression weight $\omega^{reg}(d) = \frac{P(D\geq d)\cdot(\mathbb{E}[D|D\geq d]-\mathbb{E}[D])}{Var(D)}$. First, see that we can write the conditional expectation, for any $j \in \mathbb{Z}^+$:

$$\mathbb{E}[D|D \geq j] = \sum_{d\geq j} P(D = d|D \geq j)\cdot d = \sum_{d\geq j} \frac{P(D = d, D \geq j)}{P(D \geq j)}\cdot d = \sum_{d\geq j} \frac{P(D = d)}{P(D \geq j)}\cdot d. \tag{S6}$$

Next, from Equation (S6), note that the numerator of the ratio before is equivalent to:

$$\begin{aligned}
\sum_{j=d}^{\bar{N}} w_j &= \sum_{j=d}^{\bar{N}} P(D = j)\cdot(j - \mathbb{E}[D]) = \sum_{j=d}^{\bar{N}} P(D = j)\cdot j - \sum_{j=d}^{\bar{N}} P(D = j)\cdot \mathbb{E}[D] \\
&= P(D \geq d)\cdot \mathbb{E}[D|D \geq d] - \sum_{j=d}^{\bar{N}} P(D = d)\cdot \mathbb{E}[D] \\
&= P(D \geq d)\cdot \mathbb{E}[D|D \geq d] - P(D \geq d)\cdot \mathbb{E}[D] \\
&= P(D \geq d)\cdot (\mathbb{E}[D|D \geq d] - \mathbb{E}[D]),
\end{aligned} \tag{S7}$$

where the first equality is by algebra; the second equality is by Equation (S6); the third equality holds by the complement in Kolmogorov's axioms of probability. Lastly, we know that by (SA.7) the denominator in the ratio is $Var(D)$. Both (S6) and (S7) imply that the regression weight can be written as:

$$\omega^{reg}(d) = \frac{\sum_{j=d}^{\bar{N}} w_j}{\left(\sum_{d=1}^{\bar{N}} d\cdot w_d\right)} = \frac{P(D \geq d)\cdot(\mathbb{E}[D|D \geq d] - \mathbb{E}[D])}{Var(D)},$$

as desired. □

Lemma SA.9 (Non-negative Weights with Marginal Primitives)**.** *The weights in Proposition B.1 from a regression involving marginal building blocks are non-negative.*

Proof of Lemma SA.9. First, the aggregate variable D is discrete with positive support no matter if the support of the discrete sub-treatments are binary or multivalued. Next, note that the conditional probabilities that make up the weights under causal-identifying assumptions at the disaggregated level are non-negative by axioms of probability. Hence, the only term to study non-negativity is for the $\omega^{reg}(d)$ term, $\forall d \in \mathcal{D} := \{1, \ldots, \bar{N}\}$, where $\bar{N} \in \mathbb{Z}^+$ is the largest possible value in the support of the aggregated treatment variable D.

For any $d \in \mathcal{D}_{>0}$, recall that the regression weight $\omega^{reg}(d)$ can be written as: $\omega^{reg}(d) = \frac{\sum_{j=d}^{K} w_j}{\sum_{d=1}^{\bar{N}} d \cdot w_d}$, where $w_j = P(D = j) \cdot (d - \mathbb{E}[D])$. We break the proof into two parts. First we show that the denominator is positive; and then we prove that the numerator of $\omega^{reg}(d)$ is non-negative.

1. *Denominator:* See that the denominator of $\omega^{reg}(d)$ is the variance of D by Lemma SA.7, which is positive for non-degenerate D. Thus, it suffices to show that the numerator is non-negative.

2. *Numerator:* We show that the numerator is non-negative. Take any $\bar{d} \in \{1, \ldots, \bar{N}\}$ from the support of the aggregated treatment D. Assume for the sake of contradiction that the numerator is negative. This implies the following:

$$
\begin{aligned}
\sum_{j=\bar{d}}^{\bar{N}} w_j < 0 &\iff \sum_{j=\bar{d}}^{\bar{N}} P(D = j) \cdot (j - \mathbb{E}[D]) < 0 \\
&\iff \sum_{d=1}^{\bar{N}} P(D = d) \cdot (d - \mathbb{E}[D]) - \sum_{d=1}^{\bar{d}-1} P(D = d) \cdot (d - \mathbb{E}[D]) < 0 \\
&\iff \sum_{d=1}^{\bar{N}} \cdot P(D = d) \cdot d - \sum_{d=1}^{\bar{N}} P(D = d) \cdot \mathbb{E}[D] - \sum_{d=1}^{\bar{d}-1} P(D = d) \cdot d + \sum_{d=1}^{\bar{d}-1} P(D = d) \cdot \mathbb{E}[D] < 0 \\
&\iff \mathbb{E}[D] - \mathbb{E}[D] \cdot \sum_{d=1}^{\bar{N}} P(D = d) + \sum_{d=1}^{\bar{d}-1} P(D = d) \cdot \mathbb{E}[D] < \sum_{d=1}^{\bar{d}-1} P(D = d) \cdot d \\
&\iff \mathbb{E}[D] \left(1 - \sum_{d=1}^{\bar{N}} P(D = d) \right) + \sum_{d=1}^{\bar{d}-1} P(D = d) \cdot \mathbb{E}[D] < \sum_{d=1}^{\bar{d}-1} P(D = d) \cdot d \\
&\iff \mathbb{E}[D] \cdot P(D = 0) + \sum_{d=1}^{\bar{d}-1} P(D = d) \cdot \mathbb{E}[D] < \sum_{d=1}^{\bar{d}-1} P(D = d) \cdot d \\
&\iff \mathbb{E}[D] \left(\sum_{d=0}^{\bar{d}-1} P(D = d) \right) < \sum_{d=0}^{\bar{d}-1} P(D = d) \cdot d \\
&\iff \mathbb{E}[D] < \frac{\sum_{d=0}^{\bar{d}-1} P(D = d) \cdot d}{\sum_{d=0}^{\bar{d}-1} P(D = d)},
\end{aligned}
\tag{S8}
$$

where the first line is by definition of w_j; the second line is by the complement of the previous sum; the third line is by the distributive property in algebra; the fourth line is by the definition of expectation and some algebra; the fifth line is by the complement axiom of probability and algebra; the sixth line is by the definition of probability at $D = 0$; the seventh line is by combining the terms on the LHS of the inequality; and the eighth line is by division, given non-degeneracy of the distribution of D.

If the relationship in (S8) holds, then any complete set of probabilities for a non-negative, discrete random variable will satisfy (S8). For this reason, let D be discrete uniform in probability. That is, $\forall d \in \mathcal{D}$, $P(D = d) = 1/(\bar{N} + 1)$. From (S8), this implies the following for the LHS of the inequality:

$$\begin{aligned}
\mathbb{E}[D] &= \sum_{d=0}^{\bar{N}} d \cdot P(D = d) = 0 + 1 \cdot \left(\frac{1}{\bar{N}+1}\right) \cdots + \bar{N} \cdot \left(\frac{1}{\bar{N}+1}\right) \\
&= \left(\frac{1}{\bar{N}+1}\right) \cdot \left(1 + 2 + \cdots + \bar{N}\right) = \left(\frac{1}{\bar{N}+1}\right) \cdot \left(\frac{\bar{N}(\bar{N}+1)}{2}\right) = \left(\frac{\bar{N}}{2}\right), \qquad \text{(S9)}
\end{aligned}$$

where the first equality holds by the definition of expectation; the second equality holds by the discrete uniform random variable; the third equality follows by algebra; and the fourth equality holds by the known sum of $\bar{N}$ natural numbers. For the RHS of the inequality in (S8), we find that the numerator is:

$$\begin{aligned}
\sum_{d=0}^{\bar{d}-1} P(D = d) \cdot d &= 0 + 1 \cdot \left(\frac{1}{\bar{N}+1}\right) \cdots + (\bar{d} - 1) \cdot \left(\frac{1}{\bar{N}+1}\right) \\
&= \left(\frac{1}{\bar{N}+1}\right) \cdot \left(1 + 2 + \cdots + (\bar{d} - 1)\right) = \left(\frac{1}{\bar{N}+1}\right) \cdot \left(\frac{(\bar{d}-1)\bar{d}}{2}\right), \qquad \text{(S10)}
\end{aligned}$$

where the first equality holds by expansion of the sum; the second equality follows by algebra; and the third equality holds by applying the well known sum of $(\bar{d} - 1)$ natural numbers. And similarly, for the denominator in (S8):

$$\sum_{d=0}^{\bar{d}-1} P(D = d) = \sum_{d=0}^{\bar{d}-1} \left(\frac{1}{\bar{N}+1}\right) = \bar{d} \cdot \left(\frac{1}{\bar{N}+1}\right), \qquad \text{(S11)}$$

where the first and second equalities hold by the definition of the probabilities from a discrete uniform distribution and by the expansion of the sum, respectively. By (S10) and (S11), the RHS becomes:

$$\frac{\sum_{d=0}^{\bar{d}-1} P(D = d) \cdot d}{\sum_{d=0}^{\bar{d}-1} P(D = d)} = \frac{\left(\frac{1}{\bar{N}+1}\right) \cdot \left(\frac{(\bar{d}-1)\bar{d}}{2}\right)}{\left(\frac{\bar{d}}{\bar{N}+1}\right)} = \left(\frac{\bar{d}-1}{2}\right). \qquad \text{(S12)}$$

Thus, by (S9) and (S12), the inequality from (S8) states:

$$\mathbb{E}[D] < \frac{\sum_{d=0}^{\bar{d}-1} P(D = d) \cdot d}{\sum_{d=0}^{\bar{d}-1} P(D = d)} \iff \left(\frac{\bar{N}}{2}\right) < \left(\frac{\bar{d}-1}{2}\right) \iff \bar{N} < \bar{d} - 1 \iff \bar{N} + 1 < \bar{d}.$$

However, this is a contradiction since $\bar{d} \in \{1, \ldots, \bar{N}\}$. Therefore, it must be that the numerator $\sum_{d=d}^{\bar{N}} w_j \geq 0$.

Since the denominator is positive and the numerator is non-negative, then their ratio is non-negative. Thus, the regression weight, and hence all weights under causal identification assumptions, $\omega^{reg}(d) \geq$

0 for any $d \in \{1, \ldots, \bar{N}\}$. $\square$

Lemma SA.10 (Overall Weights from Regression with Marginal Primitives Sum to One)**.** *The weights in Proposition B.1 from a regression involving marginal building blocks sum to one.*

Proof of Lemma SA.10. This subsumes the proof that regression weights alone sum to one. For any weighting function $w(s_d, s_{d-1})$ on the marginal sub-treatment contrasts that satisfies the properties of Proposition C.3, note that the sum:

$$\begin{aligned}
\sum_{d=1}^{\bar{N}} \sum_{\substack{(s_d, s_{d-1}) \\ \in \boldsymbol{\mathcal{M}}(d)}} \omega^{reg}(d) \cdot w(s_d, s_{d-1}) &= \sum_{d=1}^{\bar{N}} \sum_{s \in \mathcal{S}_d} \omega^{reg}(d) \cdot P(S = s_d | D = d) \cdot P(S = s_{d-1} | D = d-1) \\
&= \sum_{d=1}^{\bar{N}} \sum_{\substack{(s_d, s_{d-1}) \\ \in \mathcal{M}(d)}} \frac{\sum_{j=d}^{\bar{N}} w_j}{\left(\sum_{d=1}^{\bar{N}} d \cdot w_d\right)} \cdot P(S = s_d | D = d) \cdot P(S = s_{d-1} | D = d-1) \\
&= \sum_{d=1}^{\bar{N}} \frac{\sum_{j=d}^{\bar{N}} w_j}{\left(\sum_{d=1}^{\bar{N}} d \cdot w_d\right)} \cdot \sum_{\substack{(s_d, s_{d-1}) \\ \in \mathcal{M}(d)}} P(S = s_d | D = d) \cdot P(S = s_{d-1} | D = d-1) \\
&= \frac{\sum_{d=1}^{\bar{N}} \sum_{j=d}^{\bar{N}} w_j}{\left(\sum_{d=1}^{\bar{N}} d \cdot w_d\right)} \cdot 1 = \frac{\sum_{d=1}^{\bar{N}} d \cdot w_d}{\sum_{d=1}^{\bar{N}} d \cdot w_d} = 1,
\end{aligned}$$

where the first and second equality hold by the definition in Proposition B.1. The third equality uses the linearity property of the summation operator. The fourth and fifth equality follow by algebra. $\square$

Lemma SA.11 (Largest Regression Weight with Marginal Primitives)**.** *The regression weights in Proposition B.1 from a regression involving marginal building blocks are increasing as d approaches* $\mathbb{E}[D]$ *and decreasing as d moves farther away from* $\mathbb{E}[D]$.

Proof of Lemma SA.11. Suppose that D is a discrete, non-negative random variable with countably finite support and mean $\mathbb{E}[D] \in \mathbb{R}^+$.[1] Denote the support of D as $\mathcal{D} := \{0, 1, \ldots, \bar{N}\}$. From the regression decomposition using the marginal-type building blocks, we saw that the regression weight for each realization of D is: $\omega^{reg}(d) = \frac{\sum_{j=d}^{\bar{N}} w_j}{Var(D)}$, where $w_j = P(D = j) \cdot (j - \mathbb{E}[D])$, for all $j \geq d \in \mathcal{D}_{>0}$. Since each weight is scaled by the variance of D, a positive quantity, it will be sufficient to show that the weights are increasing/decreasing in the numerator only. Next, we break the proof into cases.

[1] The proof holds analogously for the case where support is countably infinite. Simply replace $\bar{N}$ with ∞.

1. *Case 1 (Increasing):* Take any $d \in \mathcal{D}$ such that $d < \mathbb{E}[D]$ and $d+1 \leq \mathbb{E}[D]$. This implies:

$$\sum_{j=d}^{\bar{N}} w_j = \sum_{j=d}^{\bar{N}} P(D=j) \cdot (j - \mathbb{E}[D]) = P(D=d) \cdot (d - \mathbb{E}[D]) + \sum_{j=d+1}^{\bar{N}} P(D=j) \cdot (j - \mathbb{E}[D])$$
$$= P(D=d) \cdot (d - \mathbb{E}[D]) + \sum_{j=d+1}^{\bar{N}} w_j. \tag{S13}$$

   Therefore, the difference between the weight at d and $d+1$ is:

$$\sum_{j=d}^{\bar{N}} w_j - \sum_{j=d+1}^{\bar{N}} w_j = P(D=d) \cdot (d - \mathbb{E}[D]) < 0,$$

   since probability is non-negative and by assumption $d < \mathbb{E}[D]$. Therefore, the weight at $D = d$ is smaller than the weight at $D = d+1$, $\omega^{reg}(d) < \omega^{reg}(d+1)$. This implies that the weights are increasing as d increases for values of $D < \mathbb{E}[D]$.

2. *Case 2 (Decreasing):* Take any $d \in \mathcal{D}$ such that $d > \mathbb{E}[D]$; and hence $d+1 > \mathbb{E}[D]$. From (A), this implies that the difference between the weight at d and $d+1$ is:

$$\sum_{j=d}^{\bar{N}} w_j - \sum_{j=d+1}^{\bar{N}} w_j = P(D=d) \cdot (d - \mathbb{E}[D]) > 0,$$

   since probability is non-negative and by assumption $d > \mathbb{E}[D]$. Therefore, the weight at $D = d$ is larger than the weight at $D = d+1$, $\omega^{reg}(d) > \omega^{reg}(d+1)$. This implies that the weights are decreasing as d increases for values of $D > \mathbb{E}[D]$.

3. *Case 3 (Knife-edge):* Take any $d \in \mathcal{D}$ such that $d = \mathbb{E}[D]$; and hence $d+1 > \mathbb{E}[D]$. By (S13), the difference between the weight at d and $d+1$ is:

$$\sum_{j=d}^{\bar{N}} w_j - \sum_{j=d+1}^{\bar{N}} w_j = P(D=d) \cdot (d - \mathbb{E}[D]) = 0,$$

   since probability is non-negative and by assumption $d = \mathbb{E}[D]$. From (1), we know that all weights below $\mathbb{E}[D]$ will be increasing. Since $d = \mathbb{E}[D]$, then all weights below $D = d$ must be increasing. From (2), we know all weights above $\mathbb{E}[D]$ will be decreasing. Since $\mathbb{E}[D] = d$, then all weights above $D = d$ must be decreasing. Therefore, the weight at $D = d$ is the largest when $\mathbb{E}[D] = d$.

4. *Case 4 (Fuzzy):* Take any $d \in \mathcal{D}$ such that $d < \mathbb{E}[D] < d+1$, and d and $d+1$ are the closest values on the support of $\mathcal{D}$ to $\mathbb{E}[D]$. From (S13), the difference between the weight at d and

$d+1$ is:

$$\sum_{j=d}^{\bar{N}} w_j - \sum_{j=d+1}^{\bar{N}} w_j = P(D=d) \cdot (d - \mathbb{E}[D]) < 0,$$

since probability is non-negative, and by assumption $d < \mathbb{E}[D]$. Hence, the weight at $D = d+1$ must be larger than the weight at $D = d$, $\omega^{reg}(d) < \omega^{reg}(d+1)$. Next, by re-application of (S13), see that the difference between weights at $D = d+1$ and $D = d+2$ is:

$$\sum_{j=d+1}^{\bar{N}} w_j - \sum_{j=d+2}^{\bar{N}} w_j = P(D=d+1) \cdot ((d+1) - \mathbb{E}[D]) > 0$$

since probability is non-negative, and by assumption $d+1 > \mathbb{E}[D]$. That is, the weight at $D = d+2$ must be smaller than the weight at $D = d+1$, $\omega^{reg}(d+1) > \omega^{reg}(d+2)$. Hence, $\omega^{reg}(d+1) > \omega^{reg}(d)$ and $\omega^{reg}(d+1) > \omega^{reg}(d+2)$. This implies that, no matter the absolute distance between d and $d+1$ and $\mathbb{E}[D]$, the largest weight will always go to the discrete value just above the mean, $\mathbb{E}[D]$.

Cases (1)—(4) imply that the regression assigns weights that increase as d nears $\mathbb{E}[D]$ and decrease beyond it. In case (3), the largest weight occurs exactly at the value of D that equals the mean, $E[D] = d$. In case (4), the largest weight is placed on the nearest discrete value of D just above the mean, $\lceil \mathbb{E}[D] \rceil_{\mathcal{D}}$, where $\lceil \cdot \rceil_{\mathcal{D}}$ is the ceiling function with respect to the support of D. □

Proof of Proposition B.1. We show that for any $K \in \mathbb{Z}^+$ using model (1) that the causal parameter α_1 is:

$$\begin{aligned}
\alpha_1 &= \sum_{d=1}^{\bar{N}} \omega^{reg}(d) \cdot \Big(\mathbb{E}[Y|D=d] - \mathbb{E}[Y|D=d-1] \Big) \\
&= \sum_{d=1}^{\bar{N}} \sum_{\substack{(s_d, s_{d-1}) \\ \in \mathcal{M}(d)}} \omega^{\star}(s_d, s_{d-1}) \cdot \Big(\mathbb{E}[Y|S=s_d] - \mathbb{E}[Y|S=s_{d-1}] \Big),
\end{aligned}$$

where the weights in Proposition B.1 can be extended to the sub-treatment level in the second equality under unconfoundedness (Assumption 2).

Choose any $K \in \mathbb{Z}^+$ to be the number of discrete sub-treatments with positive countably finite support, $\mathcal{S}_1 := \{0, 1, \ldots, N_1\}$, $\mathcal{S}_2 = \{0, 1, \ldots, N_2\}, \ldots, \mathcal{S}_K = \{0, 1, \ldots, N_K\}$ for $N_k \in \mathbb{Z}^+$ for all $k \in \{1, \ldots, K\}$. Denote the largest possible level of aggregated treatment as $\bar{N} := \sum_{k=1}^{K} N_k$. Our aggregated treatment variable is $D := \sum_{k=1}^{K} S_k$ with support $\mathcal{D} := \{0, 1, \ldots, \bar{N}\}$, which is countably finite.[2] Following the steps in Lemmas SA.2, SA.6 and SA.7 from above, we can write the aggregate

[2]The proof follows analogously for the countably infinite case, permitting $\bar{N} = \infty$.

parameter from the regression as:

$$\alpha_1 = \frac{Cov(Y,D)}{Var(D)} = \frac{\sum_{d=1}^{\bar{N}}\left(\sum_{j=d}^{\bar{N}} w_j\right)}{\sum_{d=1}^{\bar{N}} d \cdot w_d} \cdot \Big(\mathbb{E}[Y|D=d] - \mathbb{E}[Y|D=d-1]\Big),$$

where $\frac{\sum_{j=d}^{\bar{N}} w_j}{\left(\sum_{d=1}^{\bar{N}} d\cdot w_d\right)}$ is the regression weight, $\omega^{reg}(d)$. Next, we know by Lemma SA.8 that $\omega^{reg}(d) = \frac{\sum_{j=d}^{\bar{N}} w_j}{\left(\sum_{d=1}^{\bar{N}} d\cdot w_d\right)}$ can be written as $\frac{P(D\geq d)\cdot(\mathbb{E}[D|D\geq d]-\mathbb{E}[D])}{Var(D)}$. Thus, for any $K \in \mathbb{Z}^+$ number of non-exclusive, discrete, multivalued sub-treatments, the identified parameter under model (1) is:

$$\begin{aligned}
\alpha_1 &= \frac{\sum_{d=1}^{\bar{N}}\left(\sum_{j=d}^{\bar{N}} w_j\right)}{\sum_{d=1}^{\bar{N}} d \cdot w_d} \cdot \Big(\mathbb{E}[Y|D=d] - \mathbb{E}[Y|D=d-1]\Big) \\
&= \sum_{d=1}^{\bar{N}} \frac{P(D\geq d)\cdot(\mathbb{E}[D|D\geq d]-\mathbb{E}[D])}{Var(D)} \cdot \Big(\mathbb{E}[Y|D=d] - \mathbb{E}[Y|D=d-1]\Big) \\
&:= \sum_{d=1}^{\bar{N}} \omega^{reg}(d) \cdot \Big(\mathbb{E}[Y|D=d] - \mathbb{E}[Y|D=d-1]\Big),
\end{aligned}$$

as was to be shown, and where the regression weights satisfy the properties: (i) positive for all values of $d \in \mathcal{D}$ (Lemma SA.9); (ii) $\sum_{d=1}^{\bar{N}} \omega^{reg}(d) = 1$ (Lemma SA.10); and (iii) decreasing in distance from $\mathbb{E}[D]$ (Lemma SA.11). $\square$

SA.4 Proofs of Results from Appendix B.3

Lemma SA.12. *Under Assumptions 1, 2, 4 and 8, and if sub-treatments are observed,* $\tilde{w}^+(s_d, s_{d-1})$ *is identified and given by*

$$\tilde{w}^+(s_d, s_{d-1}) = \frac{\mathrm{P}\big(S=s_d\big|D=d\big) \times \mathrm{P}\big(S=s_{d-1}\big|D=d-1\big)}{\sum\limits_{(s'_{d-1},s'_d)\in\mathcal{M}^+(d)} \mathrm{P}\big(S=s'_d\big|D=d\big) \times \mathrm{P}\big(S=s'_{d-1}\big|D=d-1\big)}.$$

Proof of Lemma SA.12. Notice that

$$\begin{aligned}
&\mathrm{P}\big(S(d)=s_d, S(d-1)=s_{d-1}\big|D\in\{d,d-1\}\big) \\
&\quad= \mathrm{P}\big(S(d)=s_d\big|S(d-1)=s_{d-1}, D\in\{d,d-1\}\big) \times \mathrm{P}\big(S(d-1)=s_{d-1}\big|D\in\{d,d-1\}\big) \\
&\quad= \mathrm{P}\big(S(d)=s_d\big|D\in\{d,d-1\}\big) \times \mathrm{P}\big(S(d-1)=s_{d-1}\big|D\in\{d,d-1\}\big) \\
&\quad= \mathrm{P}\big(S(d)=s_d\big|D=d\big) \times \mathrm{P}\big(S(d-1)=s_{d-1}\big|D=d-1\big) \\
&\quad= \mathrm{P}\big(S=s_d\big|D=d\big) \times \mathrm{P}\big(S=s_{d-1}\big|D=d-1\big),
\end{aligned}$$

where the first equality holds by the definition of conditional probability; the second equality uses Assumption 8 on the first term in the expression; the third equality holds by applying Assumption 4 to the second term; and the last equality holds immediately since $S(d)$ is the observed sub-treatment vector given $D = d$ and $S(d-1)$ is the observed sub-treatment vector given $D = d-1$. Then, the lemma then holds by the definition of $\tilde{w}^+(s_d, s_{d-1})$ in Equation (6). $\square$

Proof of Proposition B.2. The result holds immediately from the definition of $\widetilde{\text{AMATT}}^+(d)$ and by Proposition C.1 and Lemma SA.12. $\square$

SB Supplementary Empirical Analysis

SB.1 Data Description

We present summary statistics from the full sample from Caetano, Caetano, and Nielsen (2024) based on the sub-treatment categories in Table S1. All elements of Table S1 represent means of each characteristic by sub-treatment. *Participation* denotes the number of children in the sample who participated at all in that specific enrichment activity. *Mean hours* represents the average number of hours per week that children in the survey spent on each enrichment activity. *Mean proportion of enrichment* represents the average proportion of that particular enrichment activity out of all enrichment activities performed per week. On average, children spent about 0.64 hours per week in the sample on enrichment activities, as they have been defined. Moreover, approximately 46% of the sample is ever treated through enrichment activity and 54% of the sample is untreated. We notice that most participation in enrichment is spent on lessons, and very little enrichment time is used for before & after school care programs.

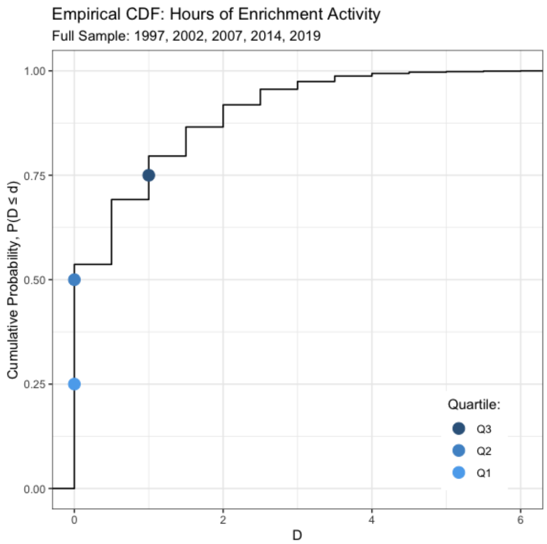
Table S1: Summary Statistics of Sub-treatments for Enrichment Activity

Sub-Treatment	Participation	Mean Hours	Mean Proportion of Enrich.
Lessons	0.31	0.25	0.50
Sports, Structured	0.19	0.19	0.25
Volunteer	0.12	0.13	0.18
Before & After School Programs	0.05	0.07	0.07

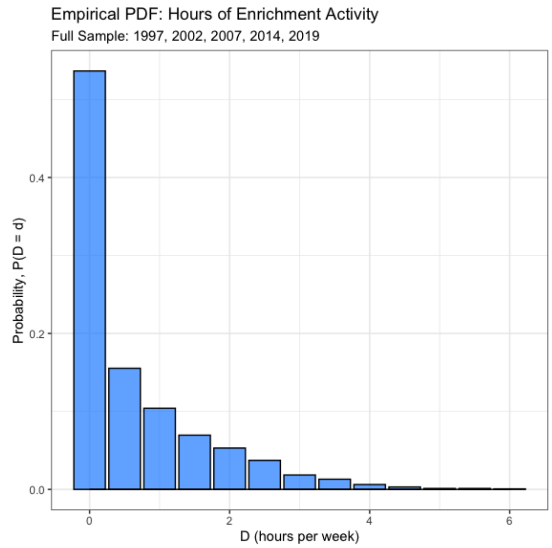
Notes: Curated data from the Childhood Development Supplement of the PSID ($N = 5,736$).

In Figure S1 below, we present the empirical CDF and PDF, respectively. We see bunching at zero enrichment hours in the distribution of aggregated treatment variable. The median total enrichment activity lies at zero hours of enrichment in a week. There are fewer and fewer children that participate in more hours of enrichment. Although seldom, some children are still observed to be participating in about an average of 4 or more hours of enrichment activities per week.

Figure S1: Empirical CDF and PDF of Total Hours of Enrichment Activities for All Children



(a) Empirical CDF.

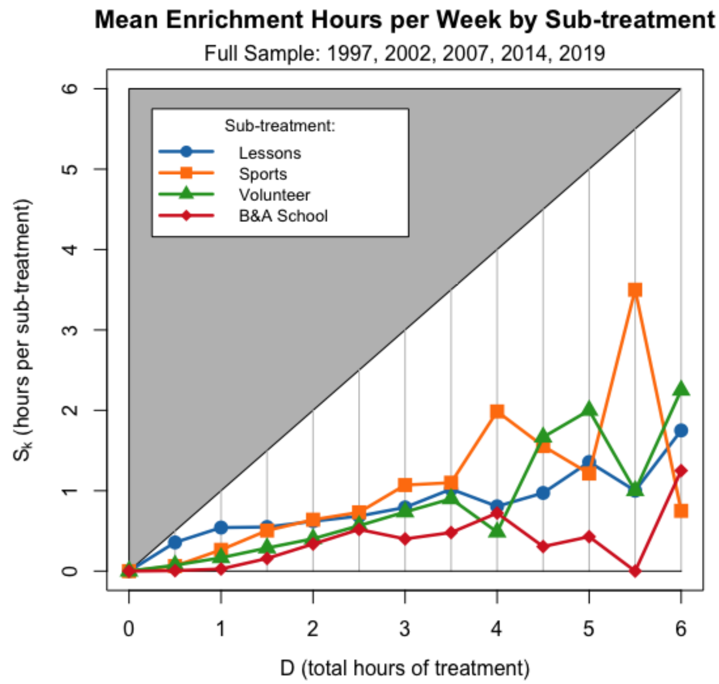


(b) Empirical PDF.

SB.2 Full Sample Analysis of Non-Cognitive Skills

This section expands the analysis from Section 5 to the full sample in Caetano, Caetano, and Nielsen (2024).

Figure S2: Average Amount of Sub-treatments across Each Level of Aggregated Treatment.



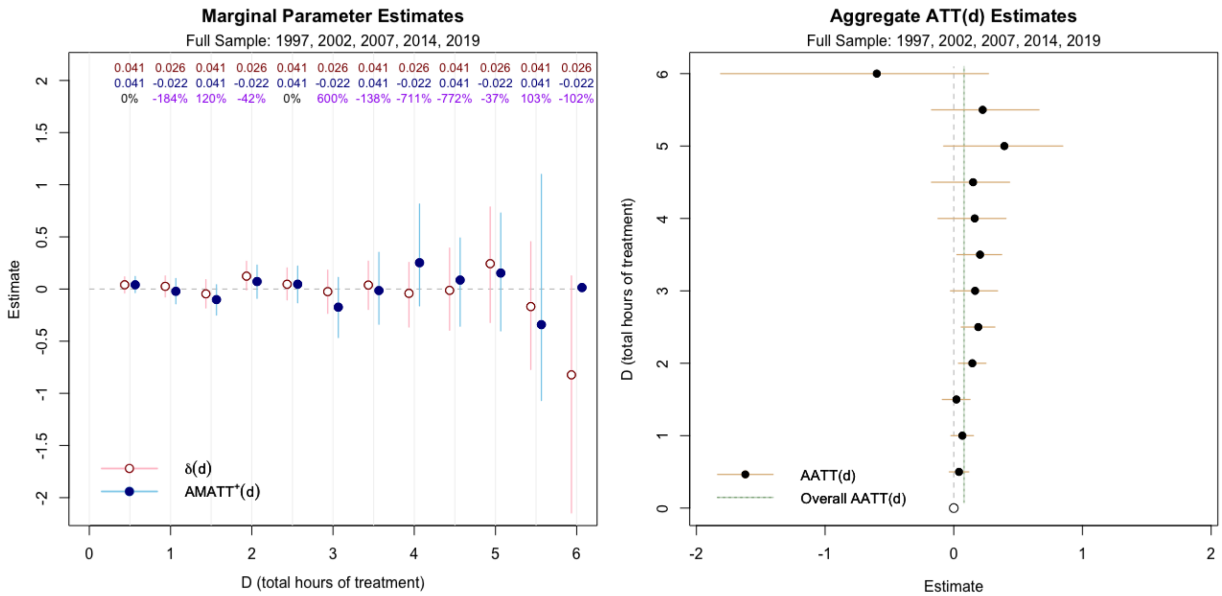
Notes: The figure displays the average amount of each sub-treatment as a function of the total amount of treatment, D.

Table S2: Overall Aggregate Parameter Estimates for Full Sample

	Parameter	Estimate	S Data	Incongruity
I. Regression				
	α_1	0.052 (0.013)		$\times$
II. Marginal				
	$\mathbb{E}[\delta(D)\|D > 0]$	0.016 (0.009)		$\times$
	$\mathbb{E}[\text{AMATT}^{+}(D)\|D > 0]$	0.002 (0.014)	$\times$	
III. Non-marginal				
	$\mathbb{E}[\text{AATT}(D)\|D > 0]$	0.080 (0.026)		
	$\mathbb{E}[\text{AATT}(D)/D\|D > 0]$	0.064 (0.032)		
Observations	5,736			

Notes: Parameter estimates of different target parameters on children's non-cognitive skills. Standard errors in parentheses obtained by bootstrap (1000 iterations). The "S Data" column indicates if sub-treatment data are required for estimation. The "Incongruity" column indicates when incongruent comparisons are present in the parameter. The estimates of $\mathbb{E}[\text{AMATT}^{+}(D)|D > 0]$ use the scaled product weights from Equation (7) in the main text.

Figure S3: d-Specific Aggregate Parameter Estimates for Full Sample



(a) Aggregate Marginal ATT.

(b) Aggregate ATT.

Notes: The figure provides estimates of the target parameters discussed in this paper for enrichment activity amounts on non-cognitive skills in children for the full sample. Panel (a) displays estimates of $\delta(d)$ and $\text{AMATT}^{+}(d)$ with 95% confidence intervals across all $D = d$. The value of each estimate, and the percent change from $\delta(d)$ to $\text{AMATT}^{+}(d)$, are listed at the top of Panel (a): $\delta(d)$—top; $\text{AMATT}^{+}(d)$—middle; and percent change—bottom. The estimates of $\text{AMATT}^{+}(d)$ use the scaled product weights from Equation (7) in the main text. Panel (b) displays estimates of $\text{AATT}(d)$ and 95% confidence intervals across all $D = d$, and an overall AATT which weights the $\text{AATT}(d)$'s by the distribution of the aggregated treatment D. At $D = 6.0$ hours of total enrichment, only a single congruent comparison is possible between $D = 5.5$ and $D = 6.0$ in the data; hence, confidence intervals are not displayed for $\text{AMATT}^{+}(d = 6)$.