

זיווג (matching) עבור הסקה סיבתית של Average Treatment Effect on Treated (ATT) במקרים של Unobserved Spatial Confounding באמצעות Distance Adjusted Propensity Score Matching (DAPSm)

עבודה בקורס הסקה סיבתית \ ד"ר דניאל נבו

תשפ"ב סמסטר א'

מגיש: אריאל וישנה, ת"ז 204149371

1.	הקדמה.....2	2.
2.	בעיית ה-Unobserved Confounders וגישת המאמר לפתרונה.....2	3.
3.	הגדרת קריטריון מרחק וה-pipeline.....4	4.
4.	השוואה מול שיטות אחרות.....5	5.
5.	ניתוח הסימולציה.....6	5.1
5.1	המודל הסימולטיבי.....6	5.2
5.2	בחינת תוצאות הסימולציה.....8	5.3
5.3	שחזור והרחבת תוצאות הסימולציה.....9	6.
6.	ניתוח תוצאות אמת.....12	7.
7.	דיון בהנחות היסוד ובתוצאות.....13	8.
8.	הרחבות אפשריות.....14	8.1
8.1	בחינת אינטראקציה בין הנתונים הנצפים לנתונים הלא-נצפים.....14	8.2
8.2	בחינת קיומם של מספר unobserved spatial confounders.....15	8.3
8.3	הרחבות אפשריות נוספות.....20	9.
9.	סיכום.....21	נספח א'
נספח א'	– זיהוי בהנחת conditional exchangeability – רקע תאורטי לפי הנלמד בקורס.....22	נספח ב'
נספח ב'	– אומדן אפקטים סיבתיים באמצעות matching – רקע תאורטי לפי הנלמד בקורס.....23	ביבליוגרפיה
ביבליוגרפיה24	

1. הקדמה

בעבודה זו נסקור המאמר "Adjusting for unmeasured spatial confounding with distance adjusted propensity score matching" מאת Georgia Papadogeorgou, Christine Choirat, and Corwin Zigler. בעבודה נציג את הבעיה העיקרית המוצגת במאמר ושיטת ההתמודדות איתה באמצעות הגדרת מדד חדש בשם distance-adjusted propensity score matching (DAPSM). לאחר מכן נשחזר את הסימולציה אשר בוצעה במאמר ונרחיב אותה, ננתח את תוצאות האמת שהוצגו במאמר, נדון בהנחות היסוד של המאמר ונציע הרחבות של המודל למקרים נוספים.

המאמר עוסק באפשרות של מימוש תהליכי אומדן סטטיסטי של אפקט סיבתי באמצעות matching במצבים שבהם יש unobserved confounders. כפי שנלמד בקורס, תהליך אמידה של אפקט סיבתי כולל שלושה שלבים מרכזיים:

1. זיהוי (identification)
2. אומדן (estimation)
3. הסקה סטטיסטית (אמידה, רווחי סמך או בדיקת השערות)

כדי להגיע לזיהוי, נדרשת אי-תלות בין התוצאה הפוטנציאלית (Y^a) לבין הסיכוי לקבל את הטיפול (A) בהנחת exchangeability ($Y^a \perp A$), אך כיוון שהנחה זו הינה מחמירה למדי בכל ניסוי שאיננו RCT ניתן להסתפק ב-conditional exchangeability על משתנים מסבירים כלשהם (X confounders) בהנחה שהם נצפים (observed confounders). על הרקע התאורטי של יכולת זיהוי ראו בנספח א'.

אם conditional exchangeability מתקיים, כלומר $Y^a \perp A|X$, לצד הנחות אחרות (positivity, SUTVA), אז ניתן לבצע תהליך זיהוי, ואז להגיע לאומדן בשיטות שונות, בין היתר באמצעות matching ובאופן ספציפי propensity score matching. מדד ה-propensity score עבור תצפית i ומשתנים מסבירים X מוגדר על ידי: $PS_i = \pi(X_i) = P(A = 1|X_i)$ ופונקציית המרחק המוגדרת באמצעותו לטובת matching הינה D_{ij} . מעתה לשם נוחות נשתמש בנוטציה שבהן השתמשנו במהלך הקורס $|\pi(x_i) - \pi(x_j)| = |PS_i - PS_j|$. במקום הנוטציה PS_i המופיעות במאמר. על הרקע התאורטי של תהליכי matching בהסקה סיבתית ראו בנספח ב'.

2. בעיית ה-Unobserved Confounders וגישת המאמר לפתרונה

ההנחה של conditional exchangeability משמעותה שאנחנו מניחים שאכן מדדנו את כל ה-covariates אשר משפיעים גם על A וגם על Y. אם זה לא המצב ויש unobserved confounders אז אנחנו בבעיה כי אז התניה על X היא לא מספיקה בשביל זיהוי. במינוח שבו הדבר מוצג במאמר, אם נסמן ב-C את כלל ה-covariates של Y ו-A, הרי ש- $C = X \cup U$ כאשר X הינם הנתונים הנצפים, מתוכם ניתן לבחור (באמצעות קריטריונים גרפיים של d-separation) את תת-הקבוצה הנצפית המינימלית שהתניה עליה תאפשר conditional exchangeability (אם הדבר מתקיים), ואילו U היא קבוצת הנתונים הלא-נצפית. אם $U \neq \emptyset$ ולא קיימת תת-קבוצה של X שמאפשרת זיהוי, אזי פירוש הדבר שישנם unobserved confounders אשר מונעים מאתנו להגיע לזיהוי. במונחים גרפיים זה אומר שעדיין יש מסלול פתוח (מסלול קורלטיבי ולא מסלול

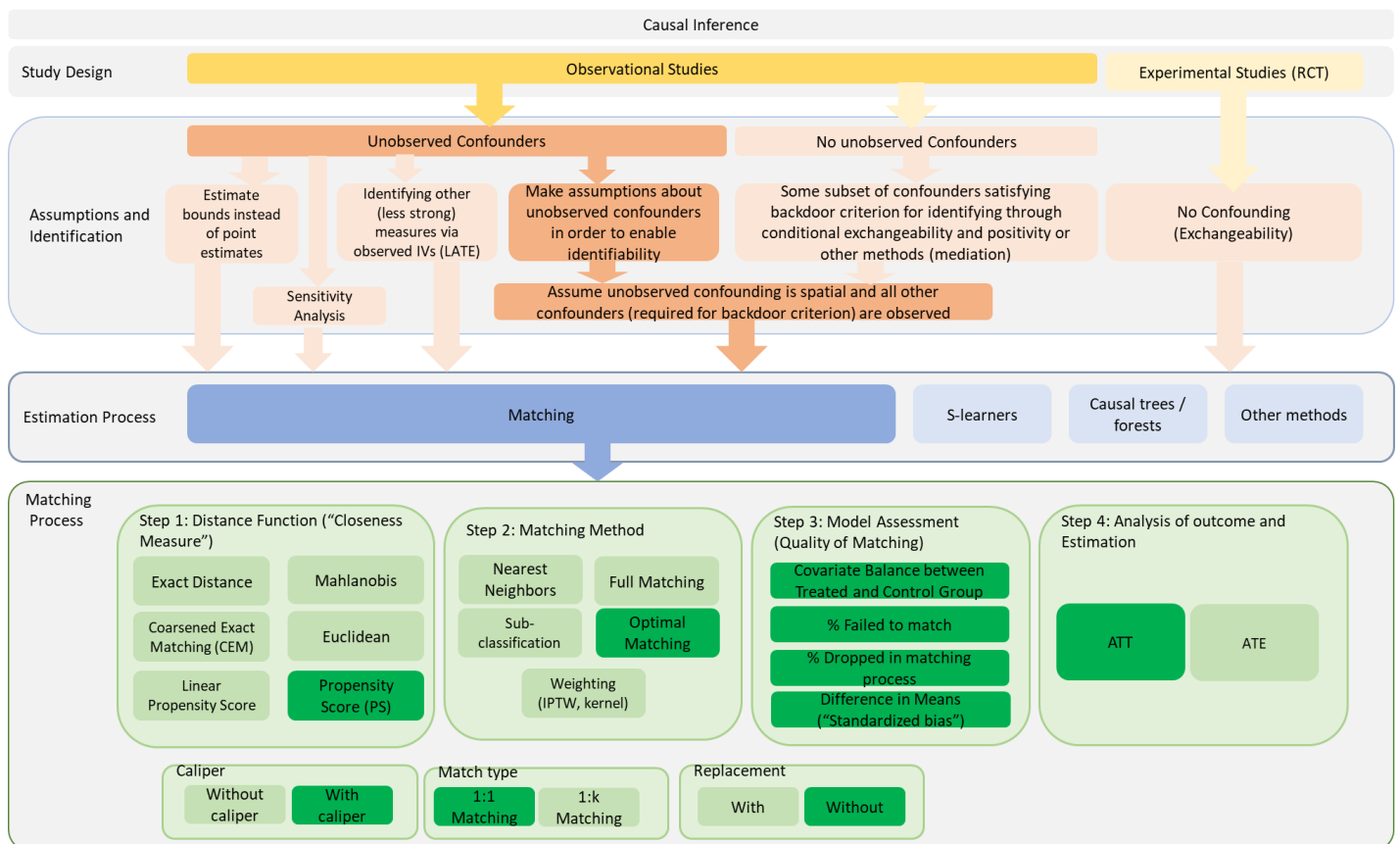
סיבתי) מ- A ל- Y ולכן אנחנו לא יכולים למדוד את האפקט הסיבתי. אינטואיטיבית, זה אומר שיש איזשהו גורם מסביר אחר גם ל- A וגם ל- Y שהוא מה שיוצר את הקשר הקורלטיבי ביניהם בלי שזה יהיה בכלל סיבה (A) ותוצאה (Y).

את הגישות השונות להתמודדות עם הבעיה של unobserved confounding ניתן לחלק לכמה וכמה קטגוריות, לצורך פשטות וחיידוד הרעיון במאמר נחלק את הגישות לשלוש שיטות עיקריות (כלל הגישות ותרשים הזרימה המחשבתי הקודם להן מוצגים בתרשים מספר 1):

גישה א': להסתפק ב-observed confounders ("הגישה הנאיבית" \ "הנועזת") – גישה שניתן לראות אותה כנאיבית במקצת היא להניח שניתן להגיע לאומדן למרות קיומם של unobserved confounders ולקיים את תהליך ההסקה למרות זאת. אפשר לתת נימוקים שונים להגיון בשיטה הזאת, בבסיסם תהיה הצדקה כלשהי של המודל מבחינת domain expertise, כאשר עיקר הטענה היא שה- unobserved confounders הם זניחים, או שקיימת הטיה מסוימת אך אנחנו מקווים שהטיה זו קטנה יחסית. מבחינה זו, זאת גם גישה נועזת כיוון שהיא חובקת בתוכה הנחה שאנחנו סומכים מאוד על המודל שלנו ועל כך שלא קיימים unobserved confounders שלא חשבנו עליהם במודל ושעלולים לשבש את התוצאות אם קיימים.

גישה ב': להסתפק במדדים אחרים ("הגישה הצנועה") להסתפק באומדן מצומצם יותר גישה אחרת שלא רואים במאמר היא בעצם לוותר על נסיון לאמוד את ה-ATE ולנסות להסתפק באומדנים יותר 'צנועים', למשל על ידי שימוש ב-Instrumental Variable שאנחנו מניחים שהוא לא מושפע מ- U כדי למדוד אומדנים כגון $LATE/CACE$. שיטות אחרות הן לוותר על אומדן נקודתי (point estimate) ובמקום זאת לנסות לאמוד רווחי סמך (bounds). הצד השווה של כל השיטות האלו הוא שאנחנו מוותרים על האומדן של האפקט הסיבתי עצמו.

גישה ג': להניח עוד הנחות על U ("גישת המאמר") – הגישה שהמאמר נוקט בה היא הגישה שאומרת שאמנם אנחנו לא יודעים למדוד את ה-unobserved confounder, אבל אנחנו יכולים להניח עליו משהו ובאופן זה לנסות לעשות לו adjustment למרות שלא ראינו אותו בפועל. היא מניחה שאנחנו יודעים להסביר את האופי של U למרות שאיננו נצפה, כלומר אנחנו לא מקבלים מודל שבו אומרים ש- U זו מין תופעה כלשהי שאין לנו דרך להסביר אותה. ספציפית במאמר אנחנו מניחים ש- U הוא **unmeasured spatial confounder**, כלומר הוא תלוי בממד גאוגרפי ושהממד הגאוגרפי הזה הוא משתנה מסביר גם של הטיפול וגם של התוצאה הפוטנציאלית. זו הנחה די חזקה. זה אומר ש- X מכיל את כל הדברים האחרים וש- U רק מוגבל לאותם spatial confounders שאנחנו רוצים להבין (כלומר אנחנו משלבים כאן גם את גישה 1 שבאה ואומרת שמצאנו עבור הנתונים שניתן לצפות בהם את כל ה-confounders הרלוונטיים). הגישה של המאמר היא להניח unmeasured spatial confounding ואז לנסות לייצר פונקציית מרחק חדשה שתמצע בין ה-propensity score של ה-measured confounding לבין המרחק של ה-unmeasured spatial confounding. נשים לב כאן כמובן שהדבר הזה אפשרי רק בגלל שיש לנו הנחה לגבי האופי הרציף של U ושיש לנו את היכולת למדוד את המרחק.



תרשים 1: 'מפת דרכים' לשיטה המתודולוגית הנקוטה במאמר ביחס לשיטות אחרות בהיבטי סוג המחקר (צהוב), הנחות היסוד (אדום), שיטת האמידה הסטטיסטית (כחול) ואופן ביצוע תהליך ה-matching (ירוק). תרשים הזרימה הבסיסי הינו בדומה למודל ה-Identification-Estimation Flowchart המוצג בדי (Brady (2020) פרק 2.4). 'הבחירות המתודולוגיות' הנקוטות במאמר הינן בצבעים מודגשים ביחס לשיטות אחרות. הנחות היסוד מבטאות של שלוש הגישות המתוארות לעיל. ארבעת השלבים המובאים בתהליך ה-matching הינם בהתאם לנלמד בקורס (הרצאה 4) ול-Stuart (2010).

3. הגדרת קריטריון מרחק וה-pipeline

בהתאם לכך במאמר מוצע להשתמש בקריטריון מרחק הממצע בין ה-propensity score הידוע מה-observed confounders לבין המרחק הגאוגרפי. זאת באמצעות ההגדרה הבאה. עבור תצפיות כלשהן i, j שה-propensity score שלהן הוא $\pi(x_i), \pi(x_j)$ בהתאמה והמרחק הגאוגרפי ביניהן הוא $Dist_{ij}$, ועבור פרמטר $0 \leq w \leq 1$ כלשהו, ה'מרחק המותאם' (distance adjusted propensity score) מוגדר על ידי:

$$DAPS_{ij} = w \cdot |\pi(x_i) - \pi(x_j)| + (1 - w) \cdot Dist_{ij}$$

זהו מיצוע של המרחק הגאוגרפי ($Dist_{ij}$) לבין המרחק ב-propensity scores כפונקציה של הפרמטר w . הפרמטר w מאפשר גמישות כדי לבחור כמה אנחנו רוצים להסתמך בעיקר על הנתונים הנצפים ועל ה-propensity score שמבטא אותם, לעומת כמה אנחנו רוצים להסתמך דווקא על המרחק הגאוגרפי המבטא דווקא את ה-unobserved confounding. ניתן להבין בקלות מה משמעות w פשוט באמצעות הסתכלות על מקרי הקצה:

- כאשר $w = 1$ פירוש הדבר שאנחנו נותנים את כל המשקל למרחק של ה-PS כלומר אנחנו מתעלמים מהמרחק הגאוגרפי (ה-spatial) ולמעשה אנחנו מייצרים כאן מודל השקול לגישה א', שמניח שאפשר להסתפק ב-observed confounders כדי לייצר matching מוצלח.

- כאשר $w = 0$ פירוש הדבר שאנחנו מתעלמים לחלוטין מכל הנתונים הנצפים ומנסים לייצר צימודים רק באמצעות המרחק הגאוגרפי. הזיווגים שיווצרו לא מכילים שום מידע על ה-covariates הנצפים ולכן יבטאו קרבה גאוגרפית בלבד.

מרגע שמגדירים את פונקציית המרחק החדשה, המשך ה-pipeline כפי שמותאר במאמר הינו:

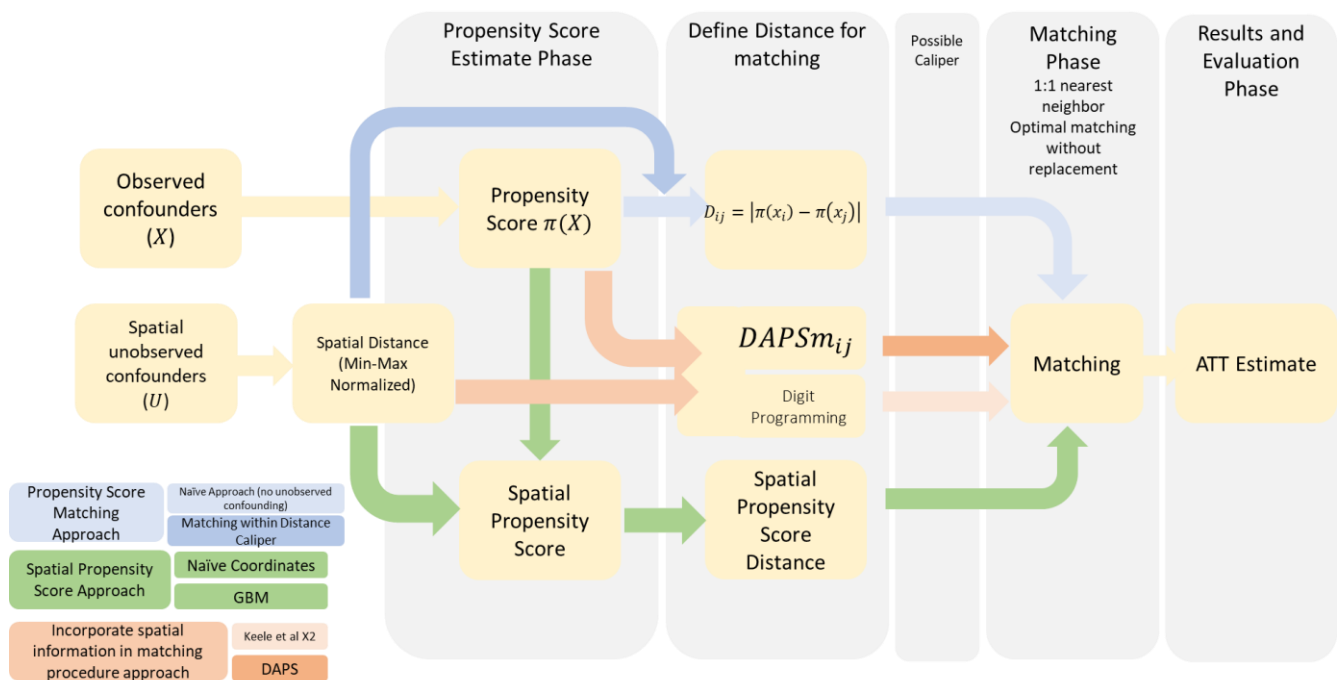
1. **בחירת w** – עקרונית אפשר לבחור את w בכל דרך רצויה, אפשר לצורך הענין לפי domain expertise לבחור w שאמור לבטא כמה אנחנו סבורים שה-unobserved confounding משפיע ביחס ל-observed confounders. הכותבים מציעים data-driven algorithm לבחירת w , זאת באמצעות תהליך איטרטיבי להבאה למינימום של ה-difference in means של ה-covariates בעקבות תהליך ה-matching.
2. **בחירת פונקציית מרחק ונרמול שלה** – כדי שלא תהיה הטיה מן הסתם בגלל שה-PS הוא בין 0 ל-1 לכן גם הערך המוחלט של המרחקים הוא בין 0 ל-1 לכן רוצים שהגדרת המרחק הגאוגרפי גם תהיה בין 0 ל-1 ולכן הבחירה הטבעית ב-min-max normalization (מרחקים לא יכולים להיות שליליים לכן זהו הנרמול המתבקש).
3. **בחירת שיטת matching** – בחירה ב-1:1 matching וב-optmatch. נציין כי בפועל אין עיסוק משמעותי בשיטות שונות של matching במאמר, כלומר לוקחים את ה-state of the art הנוכחי של matching ובעיקר מנסים לראות איך הוא מתמודד עם spatial unobserved confounding.
4. **הגדרת קליפרים** – הגדרת סף קליפר לטובת טיוב תהליך ה-matching.

4. השוואה מול שיטות אחרות

לפני שניגש לשלב ניתוח הסימולציה וניתוח נתוני האמת במאמר, נסביר את השיטות החלופיות הנסקרות במאמר לשילוב spatial unmeasured confounding שמולם המאמר משווה את השיטה של DAPSm. בסך הכל נסקרות 6 שיטות חלופיות. נקודת ההחלטה המרכזית ביותר המבחינה בין השיטות הינה שברגע שיש spatial unmeasured confounding ישנן שלוש חלופות לשילוב בתהליך ה-matching:

1. ללא שילוב – להותיר את ה-matching בהתבסס על ה-observed propensity score, כאשר ניתן לא להתייחס כלל ל-unobserved confounding או להתייחס אליו בתור caliper בלבד, כלומר להגביל זיווגים בתהליך ה-matching למקסימום מרחק גאוגרפי מסוים.
2. בתהליך אומדן ה-propensity score לשקלל את ה-spatial confounding (ולאחר מכן לבצע matching על ה-spatial propensity score בלבד).
3. בתהליך ה-matching עצמו (בנפרד מה-propensity score, כלומר ה-matching משקלל שני פרמטרים נפרדים).

כלל השיטות וההבדלים המתודולוגיים ביניהם מובאות בתרשים 2. כפי שניתן לראות, השיטה הדומה ביותר לשיטת ה-DAPSm היא שיטה שהוצעה בידי Keele et al שגם כן משקללת מידע גאוגרפי בתהליך ה-matching אך זאת ללא שימוש ב-propensity scores. כפי שעולה מהמאמר, שיטה זו מייצרת תוצאות דומות בטיבן לתוצאות המושגות בידי DAPSm.



תרשים 2: השיטות השונות הנסקרות במאמר לשילוב של *spatial unobserved confounding* בתהליך ה-*matching*. את השיטות ניתן לחלק לשלוש מתודולוגיות: התעלמות מה-*unobserved confounding* והתבססות על ה-*PS* בלבד (כחול); אומדן של *spatial propensity score* וביצוע *matching* לפיו (ירוק); אומדן נפרד של מרחק ושל *propensity score* וביצוע *matching* על בסיס שילוב של שני פרמטרים אלה (אדום). הגישה המוצעת במאמר, *DAPSm*, באדום.

5. ניתוח הסימולציה

5.1 המודל הסימולטיבי

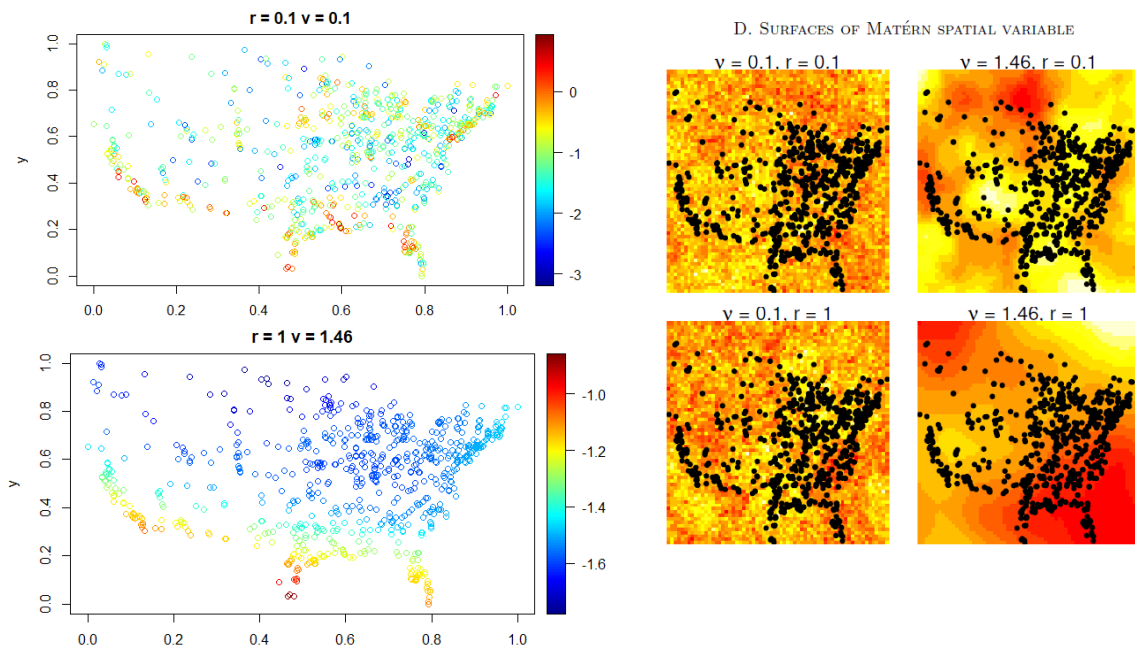
הסימולציה שבוצעה במאמר נועדה לדמות את המציאות המתוארת של *matching* תחת הימצאותם של *observed confounders* אך לצד זאת גם *unobserved spatial confounders*. בסימולציה מסמלים נתונים עבור אתרים גאוגרפיים אמיתיים שבהם קיימות תחנות כוח בארצות הברית (800 נקודות בסך הכל ב-*toyData* הראשי בחבילת *DAPSm* שמספקים כותבי המאמר). במאמר, הריצו 100 סימולציות S_1, \dots, S_{100} לכל בחירה פרמטרית r, v של פונקציית מטון (*Matérn*) המתוארת בהמשך, כאשר בכל סימולציה לכל אחת מ-800 הנקודות מגרילים ארבעה *observed confounders* המסומנים X_1, X_2, X_3, X_4 , המוגרלים מתוך התפלגות נורמלית סטנדרטית, כלומר $X_i \sim N(0, 1)$ $\forall 1 \leq i \leq 4$. לצד המשתנים הנצפים מגרילים גם *unobserved confounder* המסומן U לכל נקודת תצפית. על מנת ש- U יבטא קשר גאוגרפי כלשהו בין הנקודות, הקורלציה בין ערכי U של תצפיות שונות באותה סימולציה היא קורלציה גאוגרפית המוגרלת מתוך *Matérn Correlation* שיש לה שני פרמטרים – "חלקות" (v , smoothness) וטווח (r , range). פורמלית, השונות המשותפת של התפלגות מטון F עבור שתי נקודות במרחק h מוגדרת על ידי:

$$F(h) = \frac{1}{2^{v-1}\Gamma(v)} \left(\frac{h}{r}\right)^v K_v\left(\frac{h}{r}\right)$$

כאשר Γ היא פונקציית גאמה, K_v היא פונקציית בסל (Bessel) מסוג 2 (פונקציית נוימן) מסדר v , והמרחק h הינו לפי הגדרת פונקציית מרחק כרצוננו (אוקלידית, גאודיזית וכו'). כפי שניתן לראות בתרשים הלקוח מתוך הנספח למאמר, התפלגות מטון יכולה לסמלך נתונים גאוגרפיים תוך שקלול של חלקות, כלומר מה מידת

ה'רעש' הצפויה לעומת רציפות, וטווח, כלומר מה מידת ההשפעה שאנחנו צופים לערכים בנקודות ידועות על הסביבה שלהן. חלקות גבוהה וטווח גדול ייצרו תמונה חלקה מאוד ומכאן ש- $unobserved\ spatial$ confounding ישפיע רבות על הסביבה; חלקות נמוכה וטווח קטן מייצרים מפה יותר fuzzy שבה ההשפעה של מדידה בנקודה כלשהי של U על נקודות שכנות היא נמוכה יותר.

לאחר הגרלת U מנרמלים אותו כך שיתפלג נורמלית סטנדרטית גם כן. הקורלציה בין המשתנים הנצפים ובינם לבין המשתנה הלא נצפה U היא 0. פורמלית, $Cov(X_i, X_j) = 0, cov(X_i, U) = 0, \forall 1 \leq i \neq j \leq 4$.



תרשים 3: סימולציה של פונקציית מטון עם פרמטרים שונים r, v עבור 800 הנקודות בסימולציה. הנקודות מייצגות פריסה גאוגרפית אמיתית של תחנות כוח בארצות הברית. לקוח מתוך הנספח למאמר, חלק D (ימין); שחזור הסימולציה שביצענו באמצעות חבילת DAPSm של כותבי המאמר, עבור ערכי הקיצון $r = 0.1, v = 0.1$ ו- $r = 1, v = 1.46$ (משמאל). ככל ש- r ו- v גדולים יותר אנחנו נצפה לראות קורלציה חזקה יותר של נקודות במרחק קטן.

הטיפולים הם בינאריים $Z \in \{0,1\}$ ומוקצים לתצפיות באופן רנדומלי באמצעות הנוסחה:

$$\text{logit}(P(Z = 1)) = -0.85 + 0.1X_1 + 0.2X_2 - 0.1X_3 - 0.1X_4 + 0.3U \Rightarrow$$

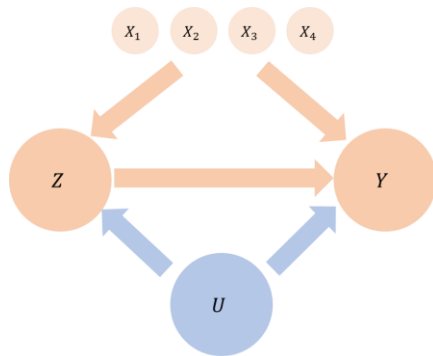
$$P(Z) = f_Z(X_1, X_2, X_3, X_4, U, \beta_Z) = \text{expit}(-0.85 + 0.1X_1 + 0.2X_2 - 0.1X_3 - 0.1X_4 + 0.3U)$$

בסך הכל מתוך הנוסחה יוצא שכ-30% מהתצפיות מקבלות את הטיפול ($Z = 1$) והיתר לא מקבלות את הטיפול ($Z = 0$). אחוז זה של תצפיות מטופלות התקבל גם אצלנו בשחזור הסימולציה.

לבסוף, התוצאה Y מוגדרת לפי:

$$Y = f_Y(Z, X_1, X_2, X_3, X_4, U, \epsilon_Y) = Z + 0.55X_1 + 0.21X_2 + 1.17X_3 - 0.11X_4 + 3U + \epsilon_Y$$

כאשר ϵ_Y רעש מתפלג נורמלית סטנדרטית.



Original model from the simulation in the paper

$$\begin{aligned} X_1 &= f_{X_1}(\epsilon_{X_1}) = \epsilon_{X_1} \sim N(0,1) \\ X_2 &= f_{X_2}(\epsilon_{X_2}) = \epsilon_{X_2} \sim N(0,1) \\ X_3 &= f_{X_3}(\epsilon_{X_3}) = \epsilon_{X_3} \sim N(0,1) \\ X_4 &= f_{X_4}(\epsilon_{X_4}) = \epsilon_{X_4} \sim N(0,1) \\ U &= f_U(h, r, v) = \frac{1}{2^{v-1}\Gamma(v)} \left(\frac{h}{r}\right)^v K_v\left(\frac{h}{r}\right) \\ Z &= f_Z(X_1, X_2, X_3, X_4, U, \beta) \\ &= \text{expit}(\beta + \alpha_{Z1}X_1 + \alpha_{Z2}X_2 + \alpha_{Z3}X_3 + \alpha_{Z4}X_4 + \alpha_{ZU}U) \\ &= \text{expit}(-0.85 + 0.1X_1 + 0.2X_2 - 0.1X_3 - 0.1X_4 + 0.3U) \\ Y &= f_Y(Z, X_1, X_2, X_3, X_4, U, \epsilon_Y) = \\ &= \alpha_{YZ}Z + \alpha_{Y1}X_1 + \alpha_{Y2}X_2 + \alpha_{Y3}X_3 + \alpha_{Y4}X_4 + \alpha_{YU}U + \epsilon_Y \\ &= Z + 0.55X_1 + 0.21X_2 + 1.17X_3 - 0.11X_4 + 3U + \epsilon_Y \end{aligned}$$

Full SEM for simulation paper

תרשים 4: גרף המתאר את המודל שמוצע בסימולציה למאמר (משמאל) לצד פירוט מלא של מערכת המשוואות SEM (מימין)

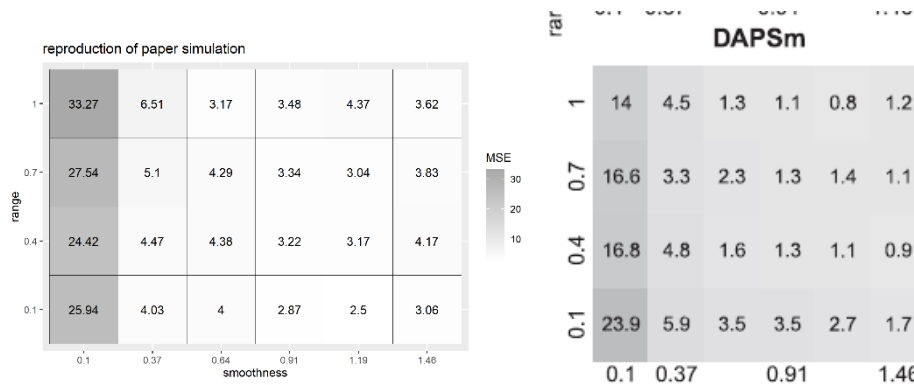
5.2 בחינת תוצאות הסימולציה

המדדה של התוצאות של הסימולציה נעשית בשיטה הבאה – עבור כל בחירה של r, v , בכל איטרציה S_1, \dots, S_{100} (הגרלה של המשתנים (X_1, X_2, X_3, X_4, U) מגדירים Gold Propensity Score (*GoldPS*) שהוא זה שמייצר את ההתאמות הטובות ביותר מבין PS תחת צפייה ב- U (כלומר ללא unobserved spatial confounding, היינו מתייחסים ל- U כאילו היה עוד observed covariate). כל המדדים מיושמים לכל בחירה של r, v ומחושבים כממוצע של התוצאות של המודלים ב-100 סימולציות לעומת ה-*GoldPS*. המדדים שנבחנו הינם:

1. איכות האומדן של האפקט הסיבתי - **MSE של ה-ATT** – בזיווג שנוצר מחשבים את ה-ATT (כלומר את ממוצע הפרשי ה- Y בין כל הזוגות של treated מול untreated). כיוון שראינו שקבוצת ה-treated הינה רק כ-30% מהתצפיות ואנחנו מבצעים 1:1 matching, הרי שהפרמטר הנאמד הוא האפקט הסיבתי על ה-treated. אחר כך משווים את ה-*ATT(Gold PS)* ל-*ATT(Model)* לכל אחד מהמודלים הנבחנו בסימולציה וממצעים את הריבועים כלומר מחשבים את ה-*MSE*.
2. איכות תהליך ה-matching - **%fail, %dropped, IQR, Distance** – מספר מדדים שנועדו לבחון את איכות ה-matching בפני עצמו. אחוז ה-fail משקף את אחוז התצפיות שלהן לא נמצא זיווג כלל, אחוז ה-dropped מייצג את אחוז התצפיות שלאחר שבוצע הזיווג נותרו ללא בן-זוג (למשל כאלה שהיו להם כמה זיווגים בתור אופציה), ה-IQR היא הטווח הבין-רבעוני של יחידות שנותרו ללא זיווג, ולבסוף המרחק distanceX100 הוא המרחק הגאוגרפי הממוצע בין יחידות שזווגו.
3. איכות ה-matching ביחס ל-PS – **Standardized Difference of Means** – כפי שראינו שיטה מקובלת למדוד איכות של matching היא באמצעות הפרשי ממוצעים של covariates שונים לאחר ביצוע ה-matching. בהנחה שה-matching מצליח לעשות את ה-adjustment הדרוש אז ה-difference of means המנורמל אמור להיות סביב 0. במקרה שלנו גם בודקים מה הפרשי הממוצעים של U במודלים השונים כדי לראות אם תהליכי ה-matching הצליחו לתת מענה ולנרמל גם את U למרות שהוא לא נצפה.

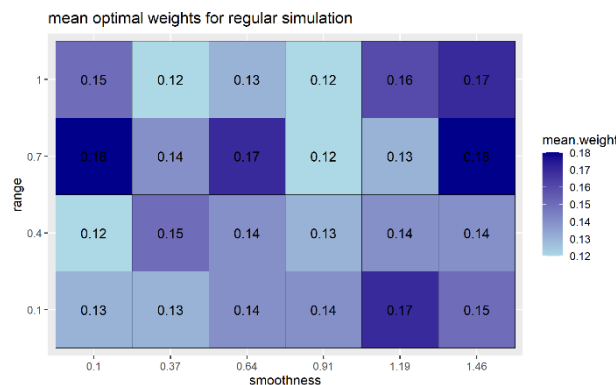
5.3 שחזור והרחבת תוצאות הסימולציה

על מנת לבחון באופן מסודר את התוצאות, ייצרנו סימולציה עצמאית שמטרתה לשחזר את עיקרי ממצאי הסימולציה במאמר, בהתבסס על חבילות שנכתבו בידי כותבי המאמר.¹ כיוון שהיינו מעוניינים לבדוק את השיטה של *DAPSm*, ביצענו סימולציה של נתונים וייצרנו matching בהתבסס על *DAPSm* והשוונו זאת למול ה-*GoldPS*, שהוא כזכור matching תחת הפיכתו של U להיות נצפה. שחזרנו את תרשים 2 מהמאמר ומדדנו עבור 4 ערכי range ו-6 ערכי smoothness את ה- MSE בין ה- ATT שחושב בהתבסס על ה-*DAPSm* לעומת ה- ATT שחושב בהתבסס על ה-*GoldPS*. מספר האיטרציות הינן 100, בהתאם למאמר המקורי. ניתן לראות את הדמיון בין התוצאות מה שמראה reproducibility של תוצאות הסימולציה.



תרשים 5: שחזור ה- MSE של סימולציה של *DAPSm* ביחס ל-*GoldPS* בהתאם לפרמטרים של המאמר. ניתן לראות כי התוצאות הינן בקירוב דומות וכי נשמר הדפוס לפיו ככל ש- r ו- v גדלים כך ה- MSE קטן

לצד זאת חישבנו גם את המשקל האופטימלי הממוצע \bar{w} לכל זוג פרמטרים r, v . נזכיר כי w נמוך פירושו שמסתמכים יותר על ה- PS כלומר פחות על U , ואילו w גבוה פירושו שמסתמכים פחות על ה- PS ולכן יותר על U . ציפינו לראות כי ככל ש- U יותר חלק ובעל טווח גדול יותר, כלומר ישנה קורלציה חזקה יותר בין שתי נקודות, הרי ש- w נמוך יותר יסקלל כיאות את U במודל. אכן ניתן לראות זאת בתוצאות שקיבלנו, מה שמראה את זה שככל שהמידע הגאוגרפי הלא נצפה הוא חזק יותר כך המודל משקלל אותו יותר בזמן ה-matching.

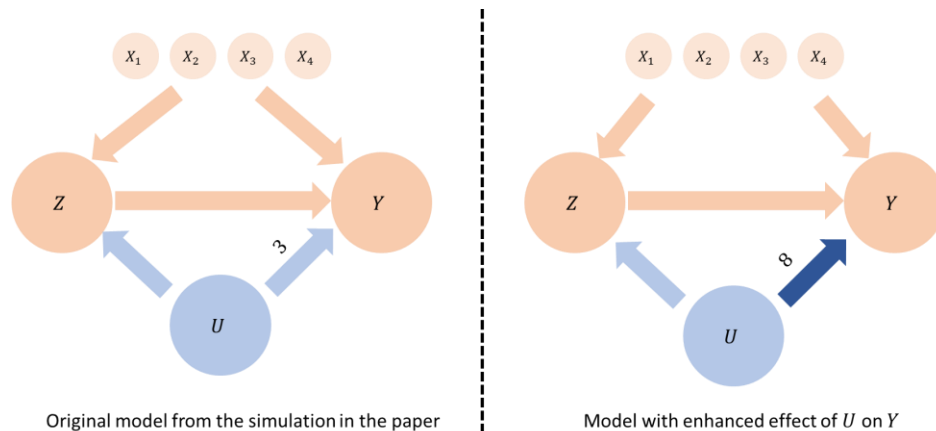


תרשים 6: משקלים w ממוצעים שחושבו לכל זוג ערכים של r, v . ניתן לראות כי ככל שהמידע הגאוגרפי חלק יותר ובעל טווח השפעה גדול יותר כך גם w גדל, כלומר משוקלל יותר ב-*DAPSm*.

¹ github.com/gpapadog/DAPSm

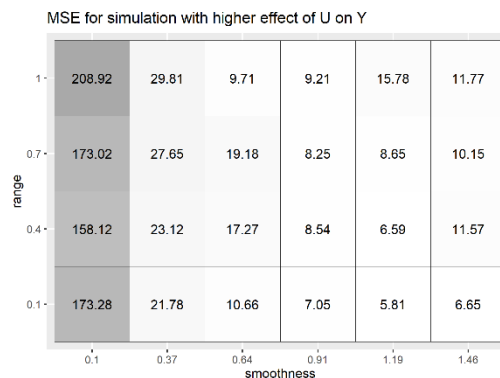
לאחר מכן ביצענו שחזור תחת הגדלת האפקט של U על Y . על ידי כך שהגדלנו את מקדם ההשפעה של U על Y מ-3 ל-8, כלומר

$$Y = Z + 0.55X_1 + 0.21X_2 + 1.17X_3 - 0.11X_4 + 8U + \epsilon$$



תרשים 7: המודל המקורי (משמאל) והמודל שבו הגדלנו את האפקט של U על Y (מימין)

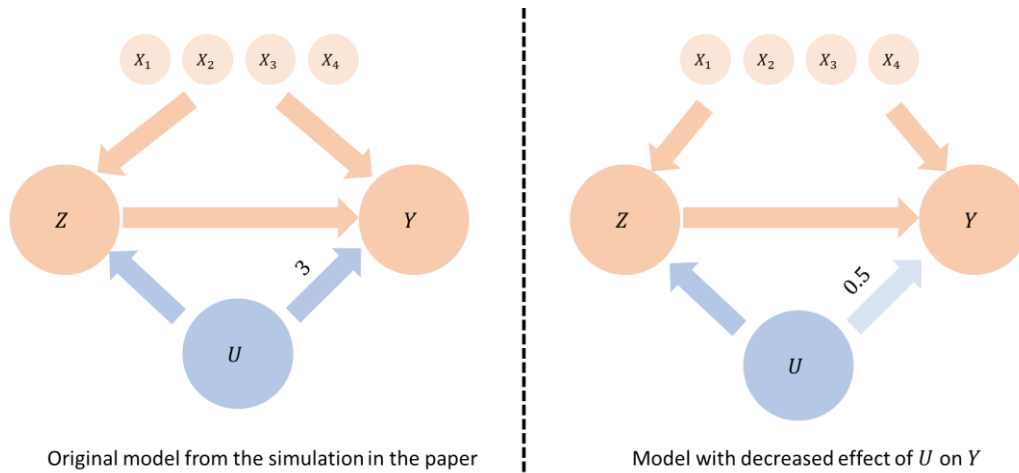
פירוש הדבר שהמידע הגאוגרפי משפיע באופן חזק הרבה יותר על התוצאה מאשר המידע הנצפה. כפי שהערכנו שיקרה, ה- MSE גדל משמעותית (בסדר גודל) עבור מידע גאוגרפי בלתי-נצפה במצבים של r ו- v נמוכים, אך גדל באופן פחות דרמטי עבור r ו- v גבוהים יותר. כיוון שלא השתנה דבר בנתונים הנצפים (השתמשנו באותם seeds ב-100 האיטרציות שביצענו) הרי שטבלת המשקלים נותרה זהה לחלוטין ביחס לשחזור הסימולציה מהמאמר ולכן איננה מוצגת כאן.



תרשים 8: ה- MSE הממוצע לרמות שונות של r , v כאשר האפקט של U על Y מוגדל. כצפוי, ההתפלגות של ה- MSE נותרה דומה (ניתן לראות כי ככל ש- r , v גדלים כך ה- MSE קטן) אך השגיאה גדולה יותר לעומת הסימולציה המקורית בגלל האפקט המוגבר

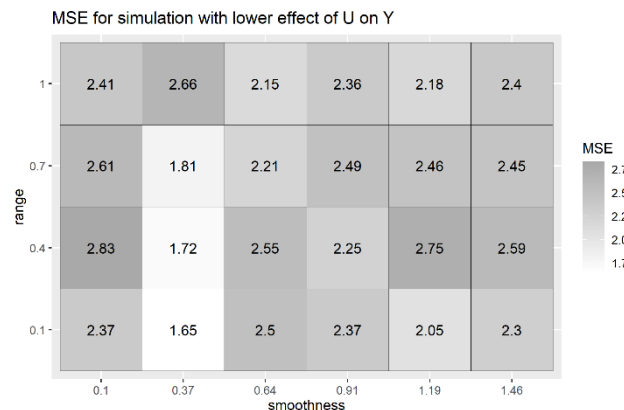
בדומה, ביצענו שחזור תחת הקטנת האפקט של U על Y על ידי הקטנת מקדם ההשפעה של U על Y מ-3 ל-0.5 כלומר:

$$Y = Z + 0.55X_1 + 0.21X_2 + 1.17X_3 - 0.11X_4 + 0.5U + \epsilon$$



תרשים 9: המודל המקורי (משמאל) והמודל שבו הקטנו את האפקט של U על Y (מימין)

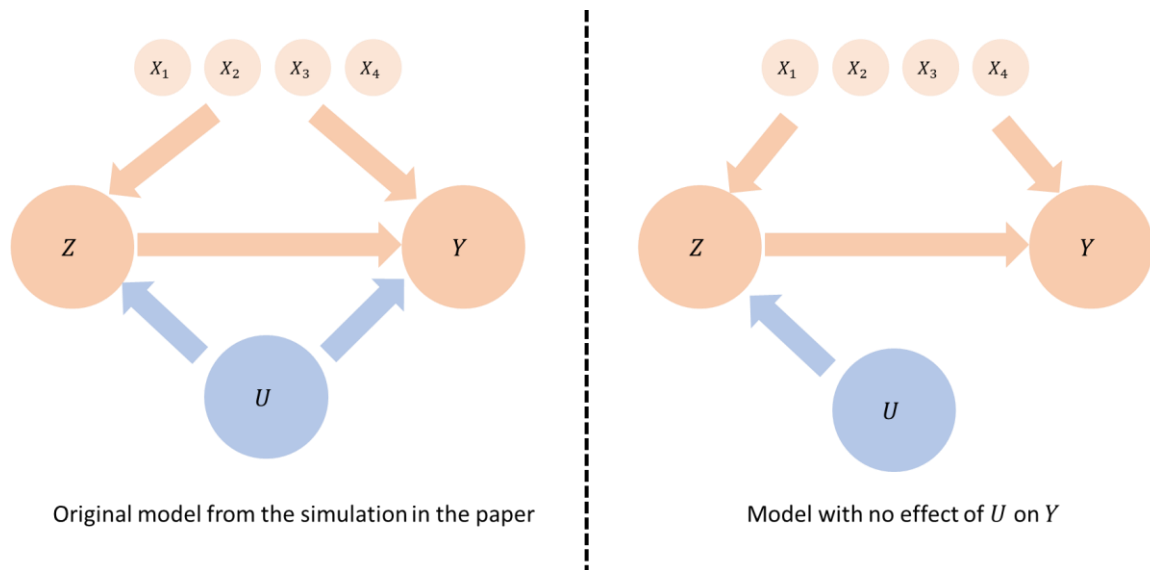
במצב כזה, כצפוי, ה- MSE קטן משמעותית עד כדי כך שלא ניתן להבחין בקשר (שראינו בסימולציות הקודמות) בין ערכי r ו- τ שונים.



תרשים 10: ה- MSE הממוצע לרמות שונות של r , τ כאשר האפקט של U על Y מוקטן. ניתן לראות כי בגלל שהאפקט קטן מאוד הרי שה- MSE קטן משמעותית ואף איננו באופן מובהק קטן יותר כאשר r , τ גדלים.

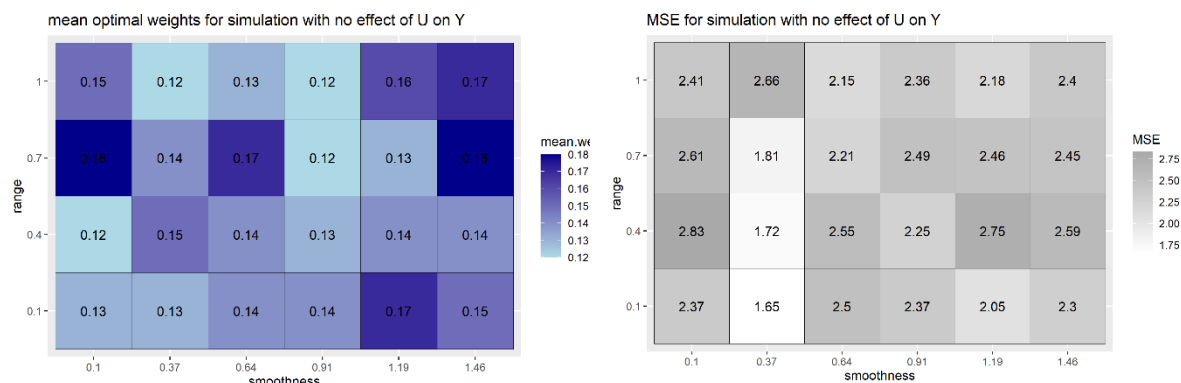
לבסוף, רצינו לבדוק מה קורה במצב שבו בכלל אין השפעה של U על Y אלא רק השפעה של U על Z , כלומר:

$$Y = Z + 0.55X_1 + 0.21X_2 + 1.17X_3 - 0.11X_4 + \epsilon$$



תרשים 11: המודל המקורי (משמאל) והמודל שבו אין השפעה של U על Y (מימין)

במקרה כזה, כפי שניתן לראות בגרף, אין בעצם מסלול קורלטיבי מ- Z ל- Y העובר דרך U ולכן בעצם אין צורך להתנות על U בשביל להגיע לאפקט הסיבתי הרצוי. ניתן לראות כי ה- w הנבחר הוא קטן מאוד בכל המודלים אך וה- MSE המתקבל נמוך מאוד בלא תלות ב- r ו- v שכן ב- $GoldPS$ אנחנו יודעים למעשה ש- U איננו משפיע בפועל על המודל.



תרשים 12: ה- MSE הממוצע לרמות שונות של r, v , כאשר אין אפקט של U על Y מוקטן. ניתן לראות כי המשקלים שניתנים קטנים מאוד (הגם שבכל זאת הם גדלים באופן יחסי ככל ש- r, v גדלים), באופן מסביר את ההתפלגות השונה של MSE ביחס לערכי r, v שונים. הדבר ממחיש את העובדה שיש סיכון בשימוש במודל $DAPSm$ כאשר בפועל אין אפקט גאוגרפי המשפיע על התוצאה הפוטנציאלית, הגם שסיכון זה הוא תחום יחסית כפי שניתן לראות מה- MSE הנמוך.

6. ניתוח תוצאות אמת

נתוני האמת שנבדקו מנתחים את ההשפעות של התקנת מתקני הפחתת פליטות של תחמוצות חנקן (nitrogen oxides, NO_x) בשם $SCR \backslash SNCR$ בתחנות כוח שונות בארצות הברית על רמות האחוזן באטמוספירה באותם אזורים (ambient ozone). בסך הכל בנתוני האמת התקבלו 473 תחנות כוח ברחי ארצות הברית, כאשר נתוני רקע דמוגרפיים באזורי התחנות ונתונים על פעילות התחנה עצמה הם ה-

observed covariates (X) והמטרה היא לאמוד את האפקט של התקנת תחנות $SCR \backslash SNCR$ (Z) על ריכוז רמות האחון בקרבת התחנה (Y). במרבית התחנות יש שילוב של מתקנים שמותקנים בהם אמצעים להפחתת פליטות ומתקנים שבהם לא מותקנים אמצעים אלה, אך בשל פשטות הפכו את הטיפול לבינארי כאשר $Z = 1$ אם מעל 50% מהמתקנים בתחנה מלווים באמצעים להפחתת פליטות. בסך הכל התקבלו 152 תחנות שבהן $Z = 1$ ו-321 תחנות שבהן $Z = 0$.

במקרה של נתוני האמת, לא ניתן לייצר *GoldPS* שבו אפשר להיעזר כדי לאמוד את טיב התוצאות. ההשוואה הבסיסית שנעשית במקרה זה היא מול *Naïve matching* אשר מתעלם מכל היבט גאוגרפי. ניתן לראות במאמר כי ה-*ATT* שמחושב בשיטה הנאיבית אכן לא מדויק, ובמקרה של השפעות על ריכוז אחון הוא אף סותר את הידע המדעי בתחום כיוון שיוצא שהתקנת מתקני הפחתת פליטות מקטינה את כמות האחון בסביבת התחנה. הדבר מחזק את הטענה שיש להשתמש ב-*distance adjustment* ואכן ה-*ATT* המחושב על ידי *DAPSm* הוא שלילי ולא חיובי. בפועל, בהשפעה על רמות האחון בסביבת התחנה האפקט הסיבתי הנאמד בידי *DAPSm* הוא קרוב ל-0, מה שמוכיח שככל הנראה אין אפקט סיבתי חזק בין התקנת מתקני $SCR \backslash SNCR$ על רמת פליטת האחון. לעומת זאת, נאמד אפקט סיבתי ממשי על פליטות תחמוצות חנקן, באופן העולה בקנה אחד עם ההבנה המדעית שתקנים $SCR \backslash SNCR$ אכן מפחיתים פליטות גזים אלה מתחנות בהם מותקנים המוצרים.

7. דיון בהנחות היסוד ובתוצאות

הבעיה העיקרית העולה מניתוח תוצאות האמת היא הבעיה בהנחה של *no interference* שהוא כזכור אחת מההנחות הנדרשות עבור *SUTVA*. אם אנחנו טוענים שהתנאים הסביבתיים משפיעים על התוצאה Y , הרי שהם אמורים להשפיע לא רק על התוצאה Y של תצפית בודדת אלא הם אמורים להשפיע כל כל התוצאות Y בסביבה מסוימת. מכאן שבמקרים רבים גם Z אמור להשפיע על Y לא רק של עצמו אלא של תצפיות אחרות. לכן ההנחה של *no interference* מתערערת במקרים רבים. רק במקרים שבהם אנחנו טוענים שהטיפול הוא חסר השפעה באופן רחב יותר מעבר למיקום הגאוגרפי הנקודתי ניתן לטעון כי הנחת ה-*no interference* איננה מתערערת, למשל במקרה שבו אנחנו רוצים לבדוק השפעה של תנאים אטמוספריים על סוגי דישונים שונים של יבולים זה הגיוני לטעון שהתנאים האטמוספריים משפיעים על איכות הגידולים אך סוג הדישון לא משפיע על התנאים האטמוספריים (ולפיכך גם לא משפיע על התנאים הסביבתיים שמשפיעים על איכות היבול הסמוך לו גאוגרפית). לעומת זאת במקרה של פליטות תחמוצות חנקן לאטמוספירה, הגיוני לטעון שיש בהחלט השפעה הדדית של וקטורי טיפולים שונים זה על זה.

להבנתנו, כותבי המאמר מודעים לאתגר הזה, לכן הם מסתפקים בלחזות את רמות האחון המקומיות זה אפקט שהם טוענים שהוא לוקאלי בטבעו. אולם אליה וקוץ בה - ברור שהסיבה שאנחנו מעוניינים לאמוד את האפקט על רמות האחון היא רק תוצר ביניים ובפועל מעניין אותנו יותר להבין דברים בהמשך השרשרת הסיבתית. למשל ההשפעה של ריכוזי אחון באטמוספירה על בריאות הציבור, על תפקוד מערכות אקולוגיות או שאלות אחרות, שבהן בבירור הגיוני שיש *interference*. ניכר שזה משהו שמעסיק את כותבי המאמר ואף

מחזר בשולי הדברים, כאשר נראה שעבודות שונות בוצעו בתחום בשנים האחרונות בדיוק כדי לנסות ולגשר על פער זה (ראו Ziegler 2021/2).

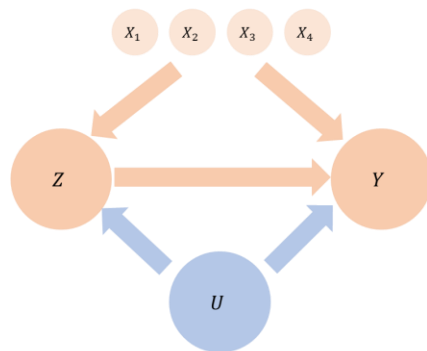
8. הרחבות אפשריות

8.1 בחינת אינטראקציה בין הנתונים הנצפים לנתונים הלא-נצפים

הסימולציה שבוצעה במאמר מניחה שיש אי-תלות בין ה-covariates הנצפים לבין U . לפי הנתון במאמר, סימולציות שבן ישנה תלות מראות תוצאות דומות אשר מביאות לידי ביטוי באופן מצומצם יותר את ההשפעה הגאוגרפית. על מנת לבדוק זאת שינינו את הסימולציה כך שתייצר תלות בין U לבין 2 מהמשתנים הנצפים - X_1 ו- X_2 . במקרה זה הגדרנו מחדש את U שנוצר מתוך $matérn$ correlation להיות:

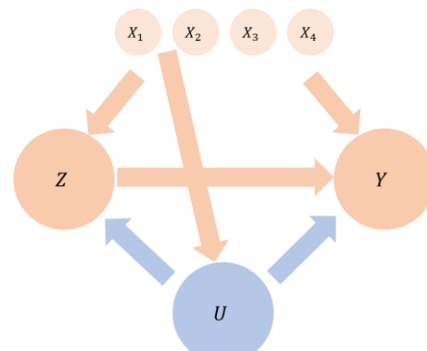
$$U := U + \epsilon_{U1}X_1 + \epsilon_{U2}X_2$$

כאשר $\epsilon_{U1} \sim N(0.25, 0.1)$ ו- $\epsilon_{U2} \sim N(0.2, 0.08)$. השינוי בערכי U מביא לכך ש- $Cov(U, X_1) \approx 0.6$ ו- $Cov(U, X_2) \approx 0.5$ אולם ממשיך להתקיים $U \perp X_3, X_4$.



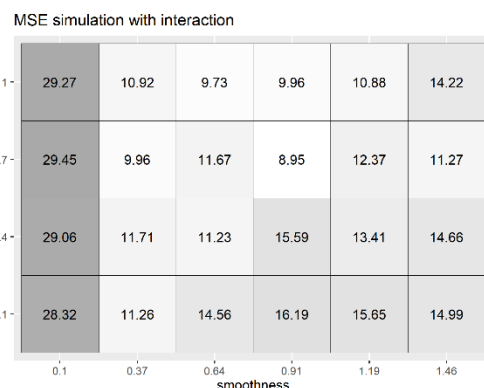
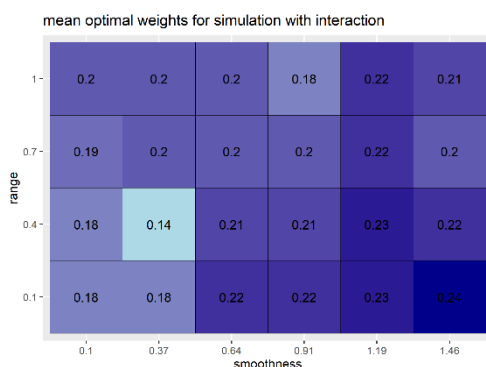
Model with interaction between observed and unobserved confounders
 $U \perp X_1, X_2, X_3, X_4$

תרשים 13: המודל המקורי (משמאל) והמודל שבו יצרנו אינטראקציה בין U ל- X_1 ו- X_2 (מימין)



Model with interactions between observed and unobserved confounders:
 $Cov(U, X_1) \neq 0, Cov(U, X_2) \neq 0$ and $U \perp X_3, X_4$

התוצאות המתקבלות הן דומות לאלו שהתקבלו במאמר על אף שה- MSE גדל פי 2 לערך עבור רוב ערכי r ו- v . יתרה על זאת, מה שניכר שמשנתה באופן מהותי הוא הקשר בין ערכים גדולים יותר של r ו- v להפחתה ב- MSE לעומת ה- $GoldPS$. ברגע שיש אינטראקציה חזקה בין המשתנים המונוטוניות המאפיינת את היחס בין ערכי r ו- v ל- MSE מצטמצמת, אם כי לא נעלמת לגמרי (בייחוד כאשר $r = 0.1$ ניתן לראות כי נותר הבדל משמעותי).

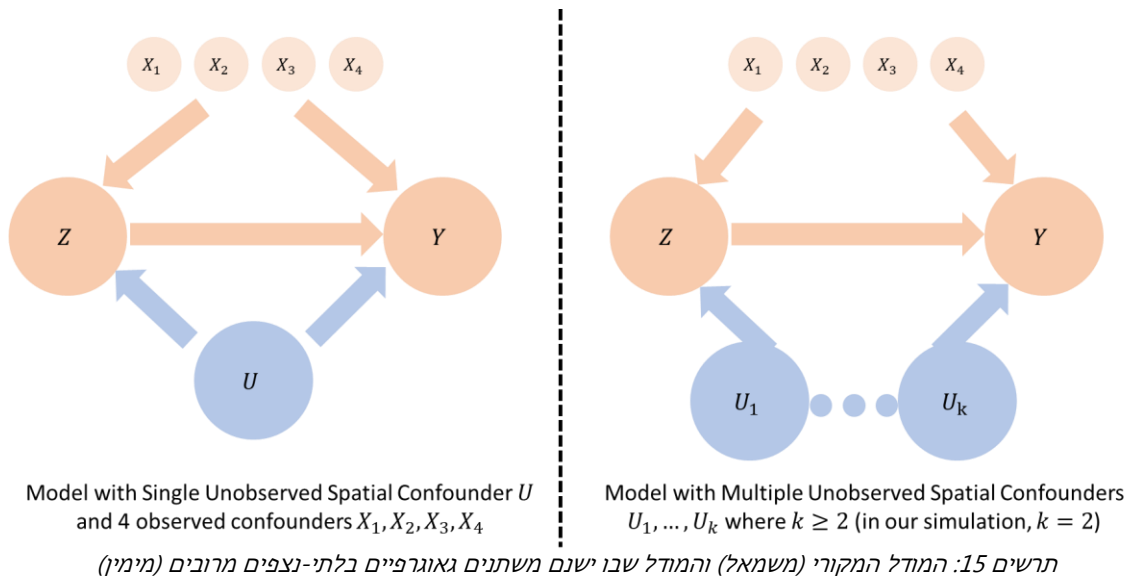


תרשים 14: MSE ממוצע ביחס ל- $GoldPS$ ומשקלים ממוצעים במצב שבו יש אינטראקציה בין U ל- X_1 ו- X_2

הסיבה העיקרית לכך להערכתנו קשורה למשקלים האופטימליים שנבחרו בידי האלגוריתם. כפי שניתן לראות התפלגות המשקלים האופטימליים על פני ערכים שונים של z ו- v היא הרבה יותר אחידה ביחס למשקלים האופטימליים שנבחרו בסימולציה הרגילה, באופן שמייצר MSE דומה על פני רוב הסימולציה.

8.2 בחינת קיומם של מספר unobserved spatial confounders

נקודה מעניינת אחת שננסה להתעכב עליה היא לשכלל את המודל למקרים שבהם החשד הוא שישנם כמה unobserved spatial confounders. כלומר אנחנו מניחים מצב שבו U הוא בעצם קבוצה של משתנים מקריים גיאוגרפיים U_1, \dots, U_k שכולם confounders. יתכן שקיימת קורלציה בין חלק מהמשתנים U_1, \dots, U_k . הרציונל מאחורי ריבוי unobserved spatial confounders הוא שבסופו של דבר קיומם של משתנים כאלו מעידים במודל מעידים על הנחה שיש איזושהי תלות גאוגרפית שאנחנו לא יודעים למדוד אך יודעים להגיד שהיא משפיעה על Y . במקרים רבים ניתן להציע מודלים שבהם ישנה תלות גאוגרפית שנובעת מכמה גורמים, כאשר מידת ההשפעה ההדדית בכל גורם משתנה. במונחי הסימולציה המתוארת לעיל, מדובר במשתנים מקריים המגיעים מפוקנציות בעלות range ו-smoothness שונים, יתכן בסדרי גודל.



דוגמאות קונקרטיות שאפשר להציע שבהן מודל מעין זה הוא רלוונטי:

- השפעה על יבולים חקלאיים – נניח שאנחנו רוצים למדוד אפקט סיבתי של דשנים שונים על טיב יבול חקלאי, לדוגמה ריכוז ויטמינים כלשהו ביבול. יתכן בהחלט שקיימת השפעה של תנאים אקלימיים גאוגרפיים בסדרי גודל גלובליים (כמו ההשפעה של תחנות הכוח המתוארת במאמר), אך במקביל לכך גם השפעה מקומית של תנאים חקלאיים כגון כמות דישון שמשפיעה על נקודות יבול סמוכות. מדובר בשני משתנים גאוגרפיים בלתי-נצפים, כל אחד בסדרי גודל אחרים, אך הם ביחד מרכיבים את U .
- השפעה על תכניות לימודים – נניח שאנחנו רוצים למדוד אפקט סיבתי של תכניות תגבור על ציונים של תלמידי בית ספר, אך אנחנו מניחים שיש confounders שאנחנו לא יודעים עליהם שקשורים לפריסה הגאוגרפית של בתי הספר. אבל במקרה זה אנחנו גם יכולים לטעון שהפריסה הגאוגרפית מורכבת מכמה משתנים מקריים נפרדים – הפריסה הגאוגרפית בתוך העיר, והפריסה הגאוגרפית הארצית.

כמובן ניתן לתת עוד דוגמאות רבות למקרים שבהם למעשה U מורכב מ- U_1, \dots, U_k . לפיכך נציע שכלול של $DAPSm$ למקרה שבו הנתונים הגאוגרפיים הם היררכיים באופיים, כלומר כל תצפית ניתן לשייך לסדרה של קטגוריות היררכיות המאפיינת אותה גאוגרפית, כאשר לכל קטגוריה גאוגרפית כזאת ניתן להגדיר צנטרואיד כלשהו ולהשוות את המרחק מצנטרואיד זה לצנטרואידים אחרים. בקטגוריה הנמוכה ביותר כל תצפית שייכת לצנטרואיד בגודל 1 שבו נמצאת התצפית עצמה בלבד. לדוגמה, אם נשווה בין בתי ספר בערים שונות שנמצאים במדינות שונות, והמודל מניח כי קיים אפקט גאוגרפי גם ברמת המרחק בין מדינות שונות, גם ברמת המרחק בין ערים שונות באותה המדינה, וגם ברמת בתי ספר שונים בתוך העיר, הרי שיש לנו שלוש קטגוריות.

פורמלית: עבור קבוצה כלשהי של תצפיות $X = \{x_1, \dots, x_n\}$ שהמיקום הגאוגרפי שלהן הוא $g(x_1), \dots, g(x_n)$ בהתאמה, נניח כי ניתן לשייך כל תצפית לסדרה של k קטגוריות גאוגרפיות היררכיות שונות $G^{(1)}, \dots, G^{(k)}$. כל קטגוריה $G^{(i)}$ (כאשר $1 \leq i \leq k$) מכילה קבוצות כמתואר להלן, ומאופיינת בשלוש תכונות:

1. הקבוצות של $G^{(i)}$ הן זרות זו לזו בזוגות כלומר לכל $A, B \in G^{(i)}$ מתקיים $A \cap B = \emptyset$.
2. איחוד כל האיברים ב- $G^{(i)}$ הוא ממצה על גבי מרחב התצפיות X כלומר $\cup_{A \in G^{(i)}} A = X$.
3. כל קבוצה ב- $G^{(i)}$ מכילה איברים כך שאיחוד של מספר קבוצות מלאות מהווה קבוצה ב- $G^{(i-1)}$. כלומר ניתן ליצור חלוקה ממצה של הקטגוריה הגבוהה יותר בהיררכיה באמצעות איחוד של קבוצות מהקטגוריה הנמוכה ביותר.

לדוגמה, נניח כי התצפיות הינן 12 בתי ספר ל-4 ערים שונות ב-2 מדינות שונות בארצות הברית, כמתואר בטבלה להלן:

מדינה $G^{(1)}$	עיר $G^{(2)}$	בית ספר $G^{(3)}$
New York	Albany	x_1
		x_2
		x_3
	Buffalo	x_4
		x_5
		x_6
Massachusetts	Boston	x_7
		x_8
		x_9
	Cambridge	x_{10}
		x_{11}
		x_{12}

טבלה 1: דוגמה ל-hierarchical spatial confounders, כל תצפית מבין x_1, \dots, x_{12} משויכת גם לקטגוריה גאוגרפית מסוג עיר וגם לקטגוריה גאוגרפית מסוג מדינה, זאת כמובן לצד הקטגוריה הגאוגרפית הבסיסית ביותר שבה כל תצפית עומדת בפני עצמה

כלומר $X = \{x_1, \dots, x_{12}\}$, מספר הקטגוריות ההיררכיות הוא 3 (בית הספר הבודד, העיר, המדינה). כאשר הקטגוריה הקטנה ביותר, $G^{(3)}$, מכילה קבוצות כמספר התצפיות, כלומר כל תצפית נמצאת בתוך קבוצה משל עצמה כלומר $G^{(3)} = \{\{x_1\}, \{x_2\}, \dots, \{x_{12}\}\}$. לאחר מכן $G^{(2)}$ היא קבוצה של ערים כלומר היא קבוצה בגודל 4 מהצורה:

$$G^{(2)} = \{\{Albany\}, \{Buffalo\}, \{Boston\}, \{Cambridge\}\} = \{\{x_1, x_2, x_3\}, \dots, \{x_{10}, x_{11}, x_{12}\}\}$$

ניתן לראות כי באמצעות איחוד של קבוצות שלמות מ- $G^{(3)}$ ניתן להגיע לקבוצות מ- $G^{(2)}$. לבסוף, $G^{(1)}$ היא קבוצת המדינה:

$$G^{(1)} = \{\{New York\}, \{Massachusetts\}\} = \{\{x_1, x_2, x_3, x_4, x_5, x_6\}, \{x_7, x_8, x_9, x_{10}, x_{11}, x_{12}\}\}$$

גם כאן ניתן לראות כי את שתי הקבוצות של $G^{(1)}$ ניתן לייצר באמצעות איחוד של קבוצות מ- $G^{(2)}$.

כעת, לכל אחת מהקטגוריות ניתן לייצר צנטרואיד עבור כל קבוצה שמייצג את המרכז הגאוגרפי של הקבוצה הנ"ל. בקטגוריה הנמוכה ביותר, בדוגמה שלנו $G^{(3)}$ הצנטרואידים הם פשוט המיקום הגאוגרפי של התצפיות עצמן כלומר אם נגדיר פורמלית פונקציית צנטרואיד $C: G^{(i)} \rightarrow |G^{(i)}|$ (פונקציה מהקטגוריה לקבוצה בגודל של הקטגוריה, כלומר כמספר הקבוצות בקטגוריה) נקבל:

$$C(G^{(3)}) = \{C(\{x_1\}), \dots, C(\{x_{12}\})\} = \{g(x_1), \dots, g(x_{12})\}$$

עבור קטגוריות גבוהות יותר בהיררכיה הצנטרואיד של כל קבוצה יוגדר בתור נקודת המרכז של הצנטרואידים של תתי-הקבוצות שלה, כלומר לדוגמה:

$$C(\{Buffalo\}) = C(g(x_1), g(x_2), g(x_3))$$

כאשר הגדרת המיצוע של צנטרואיד היא כרצוננו (לצורך הפשטות נניח אוקלידית אולם ניתן לבחור כל שיטה שהיא). כלומר אנחנו מייצרים קוארדינטות גאוגרפיות לכל קטגוריה ולכן ניתן להגדיר בהתאם סדרה של k פונקציות מרחק עבור כל קטגוריה. למשל עבור קטגוריה m שבה יש שתי קבוצות תצפיות a ו- b נוכל להגדיר:

$$Dist_{ab}^{(m)} = Dist(C(\{a\}), C(\{b\}))$$

בדומה למוצע במאמר כאשר ישנה פונקציית מרחק יחידה, גם כאן נדרש לנרמל את פונקציות המרחק באמצעות min-max normalization. עתה משיש לנו מספר פונקציות מרחק כמספר הקטגוריות, נוכל לעדכן את מדד ה- $DAPSm$ כך שישקלל בתוכו את כל פונקציות המרחק הנ"ל. נזכור כי בהגדרה המקורית מתקיים:

$$DAPSm_{ab} = w \cdot |\pi(x_a) - \pi(x_b)| + (1 - w) \cdot Dist_{ab}$$

עבור תצפית כלשהי x נגדיר $G^{(i)}(x)$ להיות הקבוצה בקטגוריה $G^{(i)}$ המכילה את x , למשל $G^{(2)}(x_1) = \{x_1, x_2, x_3\}$. עתה נגדיר סדרת מקדמים w_1, \dots, w_k שמקיימים $\sum_{i=1}^k w_i = 1 - w$. פונקציית המרחק החדשה תוגדר להיות עבור שתי תצפיות x_a, x_b :

$$DAPSm_{ab} = w \cdot |\pi(x_a) - \pi(x_b)| + \sum_{i=1}^k w_i \cdot Dist^{(i)}(C(G^{(i)}(x_a)), C(G^{(i)}(x_b)))$$

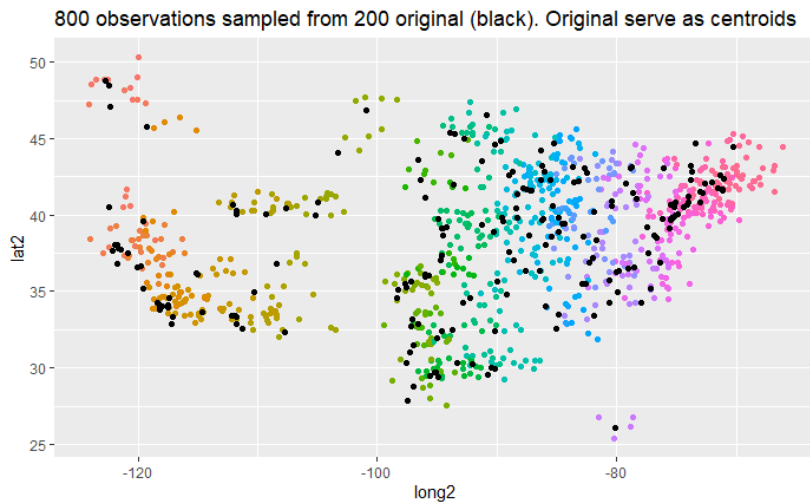
המטרה של פונקציית המרחק החדשה היא לשקלל ביחד לא רק מרחק גאוגרפי יחיד אלא סדרה של מרחקים גאוגרפיים כך שכולם יסתכמו לכדי $(1 - w)$ מתוך המרחק המשוקלל. כלומר למעשה אנחנו 'פורטים' כאן

את המרחק היחיד לכדי מיצוע (משוקלל) של k פונקציות מרחק שונות של התצפיות. האופן שבו אנחנו עושים את זה הוא לסכום את פונקציית המרחק בין הצנטרואידים של התצפיות בכל אחת מהרמות ההיררכיות השונות שלנו.

אם נחזור לדוגמה לעיל, נניח כי אנחנו נותנים משקל שווה לכל אחת מהדרגות שלנו כלומר $w_1 = w_2 = w_3 = \frac{1-w}{3}$.

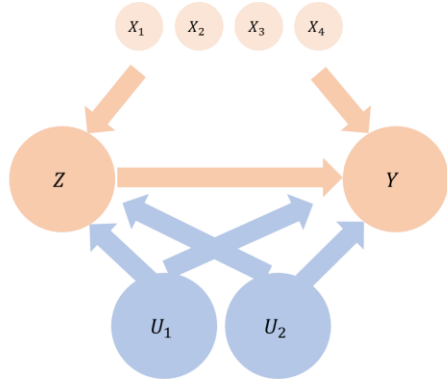
$$\begin{aligned} DAPSm_{x_1 x_7} &= w \cdot |\pi(x_1) - \pi(x_7)| + \sum_{i=1}^3 w_i \cdot Dist^{(i)} \left(C \left(G^{(i)}(x_1) \right), C \left(G^{(i)}(x_7) \right) \right) \\ &= w \cdot |\pi(x_1) - \pi(x_7)| \\ &+ \frac{1-w}{3} \left[Dist^{(1)} \left(C \left(G^{(1)}(x_1) \right), C \left(G^{(1)}(x_7) \right) \right) + Dist^{(2)} \left(C \left(G^{(2)}(x_1) \right), C \left(G^{(2)}(x_7) \right) \right) + Dist^{(3)} \left(C \left(G^{(3)}(x_1) \right), C \left(G^{(3)}(x_7) \right) \right) \right] \\ &= w \cdot |\pi(x_1) - \pi(x_7)| \\ &+ \frac{1-w}{3} \left[Dist^{(1)}(g(x_1), g(x_7)) + Dist^{(2)}(C(\{Albany\}), C(\{Boston\})) + Dist^{(3)}(C(\{New York\}), C(\{Massachusetts\})) \right] \end{aligned}$$

לטובת המחשת רעיון זה, ייצרנו סימולציה בהתבסס על הנתונים של המאמר – הפעם לקחנו את ה-dataset הקטן יותר המכיל 200 תחנות כוח ברחבי ארצות הברית (toyData2), כאשר מכל תצפית דגמנו 4 תצפיות סמוכות לה באמצעות הגרלת נקודות במרחקים מתפלגים נורמלית עם תוחלת ושונות של 0.2 סטיות תקן מתוך ה-longitude וה-latitude של הנתונים המקוריים. הצנטרואיד שהגדרנו של כל 4 נקודות היא התצפית המקורית מתוך הנתונים, כפי שמופיע בתרשים (כאן לא הגדרנו צנטרואיד לפי ממוצע אוקלידי של התצפיות האחרות אלא פשוט השתמשנו בצנטרואיד שממנו דגמנו את הנקודות האחרות).



תרשים 16: מיפוי 800 נקודות מסומלצות מתוך 200 צנטרואידים בהתבסס על מאגר toyData2 בחבילה DAPSm

הגדרנו מחדש את ה-SEM באופן הבא כך שיבטא שני משתנים גאוגרפיים בלתי נצפים (כדי שההשפעה של U_1, U_2 על Y לא תהיה גדולה מדי הקטנו את האפקטים שלהם על Y בהתאם):



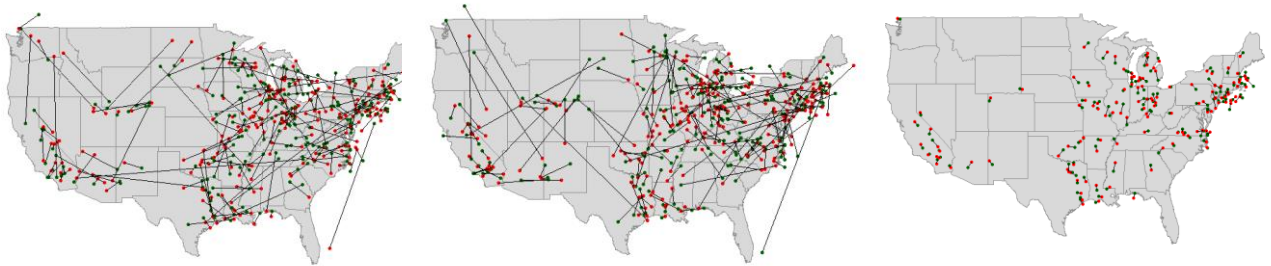
Model with 2Unobserved Spatial Confounders U_1, U_2

$$\begin{aligned}
 X_1 &= f_{X_1}(\epsilon_{X_1}) = \epsilon_{X_1} \sim N(0,1) \\
 X_2 &= f_{X_2}(\epsilon_{X_2}) = \epsilon_{X_2} \sim N(0,1) \\
 X_3 &= f_{X_3}(\epsilon_{X_3}) = \epsilon_{X_3} \sim N(0,1) \\
 X_4 &= f_{X_4}(\epsilon_{X_4}) = \epsilon_{X_4} \sim N(0,1) \\
 U_1 &= f_{U_1}(h, r_1, v_1) = \frac{1}{2^{v_1-1}\Gamma(v_1)} \left(\frac{h}{r_1}\right)^{v_1} K_{v_1}\left(\frac{h}{r_1}\right) \\
 U_2 &= f_{U_2}(h, r_2, v_2) = \frac{1}{2^{v_2-1}\Gamma(v_2)} \left(\frac{h}{r_2}\right)^{v_2} K_{v_2}\left(\frac{h}{r_2}\right) \\
 Z &= f_Z(X_1, X_2, X_3, X_4, U_1, U_2, \beta_Z) \\
 &= \text{expit}(\beta_Z + \alpha_{Z1}X_1 + \alpha_{Z2}X_2 + \alpha_{Z3}X_3 + \alpha_{Z4}X_4 + \alpha_{ZU_1}U_1 + \alpha_{ZU_2}U_2) \\
 &= \text{expit}(-0.85 + 0.1X_1 + 0.2X_2 - 0.1X_3 - 0.1X_4 + 0.3U_1 + 0.2U_2) \\
 Y &= f_Y(Z, X_1, X_2, X_3, X_4, U_1, U_2, \epsilon_Y) = \\
 &= \alpha_{YZ}Z + \alpha_{Y1}X_1 + \alpha_{Y2}X_2 + \alpha_{Y3}X_3 + \alpha_{Y4}X_4 + \alpha_{YU_1}U_1 + \alpha_{YU_2}U_2 + \epsilon_Y \\
 &= Z + 0.55X_1 + 0.21X_2 + 1.17X_3 - 0.11X_4 + 1.5U_1 + 1.1U_2 + \epsilon_Y
 \end{aligned}$$

תרשים 17: המודל הגרפי של הסימולציה שיצרנו לבדוק השפעה של מספר משתנים גאוגרפיים בלתי-נצפים (משמאל) לצד פירוט המשוואות המלא, ה-SEM (מימין)

בדקנו מקרה יחיד שבו רצינו להראות השפעה גאוגרפית שונה ברמה הלוקאלית וברמה הגלובלית, לכן הגדרנו $r_1 = 0.1, v_1 = 0.1$ עבור $G^{(1)}$ ואילו $r_2 = 1, v_2 = 1.46$ עבור $G^{(2)}$. הגדרנו משקלים שווים לשתי הקטגוריות כלומר $w_1 = w_2 = 0.5$. השתמשנו בפונקציות הנתונות בחבילת DAPSm, כאשר השינוי שביצענו היה להגדיר מחדש את פונקציית המרחק distance בפונקציה CalcDAPSWeightBalance. במקום פונקציית מרחקים רגילה (דיפולטית) השתמשנו בפונקציית מרחקים שמשקללת (לפי המשקול של w_1, w_2) את שתי הגדרות המרחקים שלנו – גם בין התצפיות וגם בין הצנטרואידים.

הרצנו קודם כל את ה-matching בהתעלם מההיררכיה ובהתייחס רק לקואורדינטות של התצפיות עצמן. באופן לא מפתיע במצב כזה (שבו בעצם אנחנו ממשקלים $w_1 = 0, w_2 = 1$) כלל הזיווגים שהתקבלו הגיעו מתצפיות סמוכות זו לזו. במצב כזה קיבלנו יחסית מעט זיווגים (כ-175) ובשגיאת טעות גבוהה יחסית ביחס ל-GoldPS. תופעה דומה אירעה במצב שבו התייחסנו רק לקלאסטרים בתהליך ה-matching. ניתן לראות כי ה-GoldPS לעומת זאת קיבל זיווגים רחוקים הרבה יותר זה מזה. לעומת זאת, כאשר משקלנו באופן הוגן יותר בין w_1 ו- w_2 גם הזיווגים הראו דפוס מרחקים הרבה יותר דומה לזיווגים של ה-GoldPS וגם מקטין את השגיאה הכללית:



תרשים 18: תוצאות ה-matching במצב שבו ישנם שני Unobserved spatial confounders. משמאל לימין: matching בהתבסס על מרחק בין תצפיות בלבד, matching בהתבסס על GoldPS, matching בהתבסס על שקלול של המרחק בין התצפיות והמרחק בין צנטרואידים. ניתן לראות שהביצועים של זיווג בהתבסס על מרחקים אבסולוטיים בלבד (בין התצפיות, ללא התחשבות בצנטרואידים של קלאסטרים) הוא מאוד לוקאלי ולא ממצה

$$sq.error(observations\ only) = sq.error(w_1 = 0, w_2 = 1) \approx 0.1$$

$$sq.error(clusters\ only) = sq.error(w_1 = 1, w_2 = 0) \approx 0.03$$

$$sq.error(combined) = sq.error(w_1 = 0.5, w_2 = 0.5) \approx 0.0028$$

נזכיר כי השקלול שעשינו כאן הוא הבסיסי ביותר ($w_1 = w_2 = 0.5$) ושיטות אופטימיזציה טובות יותר כנראה ייצרו תוצאות טובות אף יותר במקרה שבו מעריכים מראש שקיימים מספר unobserved spatial confounders. ניתן לבדוק מה יקרה במצב שבו יש גם multiple unobserved spatial confounders וגם אינטראקציות בינם לבין observed confounders. להערכתנו התוצאות יהיו דומות לסימולציה עם אינטראקציה שבוצעה בסעיף 8.1.

8.3 הרחבות אפשריות נוספות

ניתחנו לעומק שתי תוספות למודל המוצע במאמר: DAPSm עם אינטראקציה, ו-DAPSm עם מספר מרובה של unobserved spatial confounders. לצד אלה נזכיר בקצרה כמה הרחבות אפשריות נוספות:

מבינארי לרציף: לאורך הניתוח הנחנו טיפול בינארי $Z \in \{0,1\}$ ותוצאה פוטנציאלית רציפה. ניתן כמובן לשקול אפשרויות של Z קטגוריאלי או רציף, וכן בכיוון ההפוך לבדוק תוצאות בהם Y הוא קטגוריאלי או בינארי. הדבר לא צפוי להערכתנו לשנות משמעותית משהו מכל האמור לעיל מבחינת התאוריה, הסימולציה או ניתוח נתוני אמת.

שכלול תהליך ה-matching: במאמר ביצעו תהליך 1:1 matching, אולם אנחנו יכולים לשקול מעבר ל-1:k matching. הדבר רלוונטי במיוחד במקרים שבהם קבוצת המדגם של מטופלים ($Z = 1$) קטנה במיוחד. נציין כי במודל של ריבוי unobserved spatial confounders שהצענו אנחנו סבורים כי ניתן גם לשקול matching היררכי כך שיהיה אפשר לבחון matching בכל היררכיה (למשל matching בין בתי ספר שונים, אבל גם בין ערים שונות).

הרחבת הגדרת מרחק: הראינו אפשרות להרחיב את הגדרת המרחק מפונקציה יחידה לסדרת פונקציות. ברוח זו חשוב להדגיש שבמאמר התמקדו ברעיונות גאוגרפיים מתוך הנחה שה-unobserved confounders הם עם אופי רציף כפונקציה של מיקום גאוגרפי, כלומר תכונת ה-spatiality מבטאת רציפות כלשהי שאנחנו מקווים שנוכל להתמודד איתה באמצעות distance adjustment. אולם יתכן שניתן לתרגם את הרעיון של 'רציפות בלתי-נצפית' לתפיסה אבסטרקטית יותר. למשל, ניתן להסתכל על הגדרת מרחק יחסי ולא הגדרת מרחק אבסולוטי, למשל להסתכל על מרחק של בתי-ספר ממרכז עירוני וכך להשוות בין בתי-ספר עם תכונות (נצפות) דומות שגם במקרה הם במרחק דומה ממרכז עירוני כלשהו. עוד הרחבה שאפשר להעלות היא ל-time spatiality. במודלים שבהם אנחנו מעריכים שזמן הוא גם תופעה spatial באופייה, למשל במודלים הקשורים לזיהום אוויר שנמשך לאורך זמן, ניתן להרחיב את המודל כך שישקלל זמן ומרחב יחדיו (זאת, מבלי להיכנס בעובי הקורה של time-series analysis).

הוספת interference: כפי שהזכרנו בפרק 7, החיסרון הבולט המרכזי כיום במודל הוא שהוא איננו משקלל סוגיות של interference והשפעות הדדיות של תצפיות זו על זו. הוספת מודל interference איננו במוקד המאמר, אולם בכל שימוש עתידי אנחנו סבורים כי יהיה צורך להוסיף אותו למודל.

9. סיכום

בעבודה זו סקרנו מודל המציע להתמודד עם בעיית unobserved spatial confounding על ידי distance adjustment, כלומר באמצעות שקלול המרחק הגאוגרפי בין תצפיות כפתרון אפשרי לכך שלא ניתן לאמוד אפקט סיבתי במצבים שבהם confounders הם בלתי-נצפים. עמדנו על כך שלמרות שהמודל מניח הנחת יסוד חזקה מאוד לגבי ההשפעה של confounders על התוצאה הפוטנציאלית: הוא מניח שכל מה שאיננו מוסבר על ידי אפקט גאוגרפי כלשהו בלתי-נצפה (U) מוסבר על ידי נתונים נצפים (X) אשר עומדים לרשותנו בעת הניתוח. מבחינה זו, הגישה של המאמר נועזת ומתיימרת להסיק את האפקט הסיבתי ולא 'להסתפק' באפקטים סיבתיים חלקיים (ראו פרק 2).

שחזרנו את תוצאות הסימולציה שבוצעה במאמר והראינו תוצאות דומות. כאשר הגדלנו או הקטנו את ההשפעה של U על Y ראינו כי שגיאת אומדן ה- ATT ביחס ל- $GoldPS$ גדלה וקטנה בהתאמה, כפי שצפינו. הראינו כי הוספת אינטראקציה בין U ל- X מחלישה את תוצאות האמידה באמצעות $DAPSm$ אך לא מעלימה אותה לגמרי.

בניתוח נתוני האמת, עמדנו על החולשה העיקרית במודל והיא שהוא איננו משקלל interference כלומר מניח SUTVA על אף שזו הנחה לא סבירה לעתים. הדגשנו כי הדבר נכון במיוחד כאשר האפקט הסיבתי נובע מהסבר גאוגרפי (השפעה של U על Y), כאשר כיוון שזה נכון אז גם סביר ביותר שטיפולים שונים על תצפיות סמוכות ישפיעו זו על זו. הרחבת המודל למקרים של הפרת SUTVA הינם, להבנתנו, נושא משמעותי העולה ביחס ל-unobserved spatial confounding בשנים מאז כתיבת המאמר.

לבסוף, הצענו שני שכלולים מרכזיים של המודל בתצורתו הנוכחית (ללא התייחסות ל-interference): ראשית, מקרים שבהם ישנה אינטראקציה בין U לבין X ; שנית, מקרים שבהם ישנם מספר unobserved spatial confounders. הצענו מודל שמשקלל מספר קטגוריות גאוגרפיות היררכיות U_1, \dots, U_k כך שכולן ביחד ישפיעו באופן יחסי על חלקן בתוך משוואת ה- $DAPSm$. הראינו מקרה מוצלח ליישום הרחבה זו על 800 תצפיות מסומלצות הדומות באופיין למסד הנתונים של המאמר.

נספח א' – זיהוי בהנחת conditional exchangeability – רקע תאורטי לפי הנלמד בקורס

אחת הסוגיות המרכזיות בהסקה סיבתית היא שכדי להגיע ליכולת זיהוי (identification) של אפקט סיבתי (causal effect) אנחנו זקוקים להנחה של exchangeability, כלומר להניח שתוצאה פוטנציאלית כלשהי Y^a איננה תלויה בסיכוי שהתקבל טיפול A , או באופן פורמלי: הנחת exchangeability מתקיימת אם עבור מרחב טיפולים אפשריים כלשהו $a \in \mathcal{A}$ מתקיים $Y^a \perp A$. במקרה הבינארי שבו $\mathcal{A} = \{0,1\}$ נדרש שיתקיים $Y^1 \perp A$ וכן $Y^0 \perp A$. תחת exchangeability ו-SUTVA ניתן לעבור מאומדן סיבתי (counterfactual estimand) לאומדן סטטיסטי (statistical estimand) על ידי: $E(Y^a) = E(Y|A = a)$. במקרה שזה נכון ניתן לאמוד את האפקט הסיבתי הממוצע (ATE) או אומדנים סטטיסטיים דומים אחרים באמצעות:

$$ATE = E(Y^1 - Y^0) = E(Y^1) - E(Y^0) = E(Y|A = 1) - E(Y|A = 0)$$

הבעיה היא שההנחה של exchangeability היא הנחה חזקה מאוד כיוון שהיא מניחה שאין שום משפיע חיצוני (covariate) אשר קושר באופן קורלטיבי בין הטיפול לבין התוצאה באופן שאיננו סיבתי אלא קורלטיבי בלבד (correlation does not imply causation). ממסך מסוג זה (מעטה ייקרא confounder), X , מונע מאתנו להניח את ההנחה של exchangeability.

בעוד שבמחקר ניסויי (experimental studies) ניתן עדין להגיע למצב של exchangeability באמצעות רנדומיזציה בשלב הקצאת הטיפולים (RCT), הרי שבמחקרים תצפיתיים (observational studies) שבהם אין רנדומיזציה בשלב הקצאת הטיפולים זו הנחה בלתי סבירה. למשל אם בית ספר מעניק שיעורי תגבור במתמטיקה רק לילדים שמתקשים במתמטיקה, הבחירה לתת את הטיפול (שיעור התגבור) איננה בלתי-תלויה בתוצאות הפוטנציאליות (הציונים של הילדים אם היו מקבלים תגבור ואם לא היו מקבלים תגבור). בהנחה ששיעורי התגבור לא ניתנו באופן רנדומלי הרי שניסיון לאמוד את ההשפעה של שיעורי תגבור (A) על ציוני המתמטיקה (Y) הינה בהטיה שכן מראש האוכלוסיה שקיבלה שיעורי תגבור הינה לא מייצגת.

מסיבה זו, במחקרים תצפיתיים מוותרים על ההנחה המחמירה של exchangeability ומסתפקים במקום זאת בהנחה יותר סבירה של conditional exchangeability. לפי הנחה זו, כאשר קיימים covariates של האוכלוסיה, אנחנו נניח שבהינתן שאנחנו מתנים על ה-covariates (למשל אותן קבוצות גיל, אותה רמת הכנסה וכדומה) אז אנחנו כן נצפה לראות תוצאות פוטנציאליות דומות עבור הטיפול. מכאן שמגיעים להנחה ש- $Y^a \perp A|X$ ועל בסיס זאת (ביחד עם הנחת positivity) ניתן לעבור מאומדנים סיבתיים לאומדנים סטטיסטיים, כלומר להגיע לזיהוי.

הנחת ה-conditional exchangeability מאפשרת זיהוי ולפיכך אומדן סטטיסטי של האפקט הסיבתי מתוך הנחה שאנחנו יודעים מהם ה-covariates ובהתאם לכך יודעים גם להתנות עליהם. ברגע שזה המצב יש כל מיני שיטות להתניה נכונה ולאומדן של האפקט הסיבתי. מספר שיטות שהוזכרו בקורס הן IPTW, S-learners, עצים ויערות סיבתיים (causal forest, causal trees) ושיטת הזיווגים (matching) שהיא הנמצאת בליבת השיטה במאמר.

נספח ב' – אומדן אפקטים סיבתיים באמצעות matching – רקע תאורטי לפי הנלמד בקורס

בבסיסה, שיטת ה-matching נועדה לעזור לאמוד אפקט סיבתי באמצעות איתור של תצפיות אשר דומות מאוד מבחינת המאפיינים שלהן (כלומר מבחינת ה-covariates) וככל האפשר הן נבדלות אחת מהשנייה רק בדבר יחיד – אם הן קיבלו או לא קיבלו את הטיפול. בכך עוקפים את "הבעיה היסודית של הסקה סיבתית" (The Fundamental Problem of Causal Inference) באופן אלגנטי – אמנם עבור פרט יחיד לא ניתן לדעת מה היה קורה לו אילו קיבל את הטיפול או להיפך, אבל אם נמצא שני פרטים שדומים מאוד אחד לשני זה יוכל לעזור לנו בקירוב להבין מה היה האפקט הסיבתי. בהנחה שיש פחות מטופלים (treatment group) מאשר קבוצת הבקרה (control group) יש לשים לב שהאומדן הוא למעשה לא של האפקט הכולל של הטיפול (ATE) אלא של האפקט הכולל של הטיפול על המטופלים (ATT). ה-ATT מוגדר להיות:

$$ATT = E(Y^1|A = 1) - E(Y^0|A = 1)$$

אם מנסים לעשות matching באופן שבו כל ה-covariates של שתי תצפיות הם זהים לחלוטין (full matching), מגלים הרבה פעמים שהשיטה הזו מחמירה מדי באופן שמפר את הנחת ה-positivity וגם משמיט חלק גדול (אם לא את כל) מהתצפיות. הדבר מחמיר כמובן ככל שהממד של X (כלומר מספר ה-covariates שמתנים עליהם) גדל. לכן שיטה נפוצה חלופית לעשות matching היא באמצעות propensity score שהוכח כשיטה יעילה, היא נועדה לעזור לבעיות של high-dimensionality של X שלא מאפשרת matching.

$$Y^a \perp A|X \Rightarrow Y^a \perp A|P(A = 1|X) \Leftrightarrow Y^a \perp A|\pi(X)$$

כאשר ה-propensity score של X מסומן על ידי $\pi(X)$. בגלל שה-propensity score הוא balancing score אז אנחנו יכולים להשתמש בו בשביל matching כאשר הגדרת המרחק שלנו תהיה $D_{ij} = |\pi(x_i) - \pi(x_j)|$.

- Keele, L., Titunik, R, and Zubizarreta, J. (2015), "Enhancing a geographic regression discontinuity design through matching to estimate the effect of Ballot initiatives on voter turnout", *Journal of Royal Statistical Society A* 178, 223–239.
- Minansy, B. and McBratney, A. B. (2005), "The Matérn function as a general model for soil ariograms", *Geoderma* 128, 192–207.
- Neal, Brady (2020), *Introduction to Causal Inference from a Machine Learning Perspective*.
- Nevo, Daniel (2021), *Causal Inference: Lecture Notes*.
- Papadogeorgou, Georgia, Choirat, Christine and Zeigler, Corwin M. (2019), "Adjusting for unmeasured spatial confounding with distance adjusted propensity score matching", *Biostatistics* 20, 2, pp. 256–272.
- Papadogeorgou, Georgia, Choirat, Christine and Zeigler, Corwin M. (2019), "Adjusting for unmeasured spatial confounding with distance adjusted propensity score matching: Supplementary Materials", *Biostatistics* 20, 2, pp. 1-11.
- Papadogeorgou, Georgia (2019), DAPSm R package, <https://github.com/gpapadog/DAPSm.git>.
- Pearl, Judea (2009), "Causal Inference in Statistics: an Overview", *Statistics Surveys*, Vol. 3, pp. 96-146.
- Rosenbaum, Paul R., and Rubin, Donald B. (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects", *Biometrika*, Vol. 70, No. 1, pp. 41-55.
- Stuart, E. A. (2010). Matching methods for causal inference: a review and a look forward. *Statistical Science* 25, 1–21.
- Zigler, Corwin (2021), "Bipartite Inference and Air Pollution Transport", Lecture Video, *Online Causal Inference Seminar*, 15 May, 2021, <https://www.youtube.com/watch?v=PiüWu-YtIXs&t=3331s>.
- Zigler, Corwin M., and Papdogeorgou, Georgia (2021/2), "Bipartite Causal Inference with Interference", *Statistical Science*, Vol. 36, No. 1, pp. 109-123.