UNIVERSITÉ DU QUÉBEC À MONTRÉAL

ALGORITHMES DE SÉLECTION DE CONFONDANTS EN PETITE ET GRANDE DIMENSIONS: CONTEXTES D'APPLICATION CONVENTIONNELS ET POUR L'ANALYSE DE LA MÉDIATION

THÈSE

PRÉSENTÉE

COMME EXIGENCE PARTIELLE

DU DOCTORAT EN MATHÉMATIQUES

PAR

ISMAILA BALDÉ

AOÛT 2022

REMERCIEMENTS

Dans le cadre de ce travail, je tiens sincèrement à remercier ma directrice de thèse, Geneviève Lefebvre, pour sa disponibilité et son soutien.

Je remercie aussi Yi Yang, qui a accepté de collaborer avec nous dans un chapitre de cette thèse qui est publié sous forme d'article.

J'aimerais aussi remercier mon ancien encadreur de la maîtrise, Salah El Adlouni, pour ses conseils et encouragements.

Mes remerciements vont aussi à l'endroit de tous les professeurs et le personnel administratif du Département de mathématiques de l'UQAM, en particulier Karim Oualkacha. Karim est toujours disponible pour discuter de la recherche.

Enfin, je remercie toute ma famille et mes amis, pour leur présence et leur soutien.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

LISTE DES FIGURES

RÉSUMÉ

L'inférence causale cherche à déterminer les relations de causalité entre un ensemble de variables sur la base de données expérimentales ou d'observation. Alors que les études expérimentales de type randomisées sont idéales pour l'inférence causale, elles ne sont pas toujours réalisables pour des raisons logistiques, économiques ou éthiques. Il est donc particulièrement intéressant de déterminer la structure causale entre les variables sur la base de données d'observation. Dans cette thèse, nous nous intéressons aux algorithmes d'identification et de sélection des confondants basés sur les données observationnelles en trois différents projets.

D'abord, nous abordons le problème des algorithmes de sélection des confondants en grande dimension pour estimer l'effet total de l'exposition sur la réponse. Dans un article récent et populaire, Shortreed et Ertefaie (2017) ont introduit l'algorithme « outcome-adaptive lasso (OAL) » pour cibler les variables désirables pour l'estimation basée sur une méthode par scores de propension. Alors que OAL est basé sur le lasso adaptatif pour assurer la sparsité du modèle, sa capacité de tenir compte adéquatement de la colinéarité entre variables est discutable, particulièrement en grande dimension. Nous proposons le « generalized outcome-adaptive lasso (GOAL) » qui combine les forces du lasso adaptatif pour garantir la sparsité et de l'elastic net pour faire face à la colinéarité. Nous avons comparé OAL et GOAL dans les mêmes scénarios de simulation que dans Shortreed et Ertefaie (2017). Alors que les approches OAL et GOAL étaient équivalentes avec des covariables indépendantes, GOAL était plus performant que OAL dans les problèmes en grande et petite dimensions avec des covariables corrélées.

Ensuite, nous nous intéressons au sujet émergent des algorithmes de sélection des confondants pour l'analyse de médiation causale. Plus spécifiquement, nous construisons sur les algorithmes OAL et GOAL conçus pour l'estimation de l'effet total, afin de les utiliser

chacun dans une procédure algorithmique en deux étapes qui permet l'estimation des effets direct et indirect qui sous-tendent l'analyse de médiation. Nous proposons ainsi le « outcome-adaptive regularization for causal mediation analysis (MOAR) » afin de sélectionner les covariables appropriées pour l'analyse de médiation causale. Nous utilisons des données simulées et des données réelles provenant de l'étude Harvard School of Public Health College Alcohol pour illustrer la performance de MOAR. Les résultats montrent que MOAR améliore l'estimation de l'effet direct comparativement aux algorithmes standards OAL et GOAL.

Finalement, nous abordons le problème de l'identification des confondants pour l'analyse de médiation causale à partir du lasso graphique. Le lasso graphique, dit « glasso », est un algorithme très populaire et rapide qui a été introduit par Friedman, Hastie et Tibshirani (2008) pour estimer l'inverse de la matrice de variance-covariance basé sur des données normales multivariées observées. Alors que le glasso a été conçu pour estimer des graphes non dirigés, son utilité pour estimer des graphes de médiation causale est questionnable. Nous développons une nouvelle approche de régularisation, appelée « mglasso », qui permet de retrouver les liens pertinents entre variables dans un contexte de médiation causale sur la base d'une inférence par glasso. Une interface graphique basée sur l'application Shiny avec le logiciel statistique `R` est en voie de production; à terme, cette interface permettra aux utilisateurs de posséder un outil de visualisation des relations directes et indirectes probables entre l'ensemble des variables, incluant les confondants.

Mots-clés: Algorithmes d'apprentissage statistique, Analyse de médiation, Analyse statistique de données de grande dimension, Approche graphique, Inférence causale, Sélection de variables.

INTRODUCTION

L'inférence causale peut être vue comme la pierre angulaire de la découverte scientifique. Elle s'intéresse à l'étude des relations de causalité entre une variable de traitement (exposition ou intervention) et une variable réponse d'intérêt. L'être humain est naturellement familier avec les concepts fondamentaux de l'inférence causale. Hernán et Robins (2020) présentent deux exemples qui illustrent comment les humains raisonnent sur les effets causaux. Dans le premier exemple, ils considèrent un scénario avec un patient nommé Zeus. « Zeus est en attente d'une transplantation cardiaque. Le 1er janvier, il reçoit un nouveau cœur. Cinq jours plus tard, il meurt. Imaginez que nous puissions savoir, peut-être par révélation divine, que si Zeus n'avait pas reçu de transplantation cardiaque le 1er janvier il aurait été en vie cinq jours plus tard. Munis de ces informations, la plupart conviendraient que la greffe a causé la mort de Zeus. Ainsi, l'intervention de transplantation cardiaque a eu un effet causal sur la survie à cinq jours de Zeus ». Dans le second exemple, ils considèrent un autre scénario avec une autre patiente nommée Hera. « La patiente, Hera, a également reçu une transplantation cardiaque le 1er janvier. Cinq jours plus tard, elle était en vie. Imaginez que nous puissions savoir que si Héra n'avait pas reçu le cœur le 1er janvier, elle serait encore en vie cinq jours plus tard. Par conséquent, la greffe n'a pas eu d'effet causal sur la survie à cinq jours d'Héra ». Pour chacun de ces deux exemples, nous effectuons une comparaison de la réponse (décès ou non décès) lorsqu'une intervention (transplantation) est entreprise par rapport à la réponse lorsque celle-ci ne l'est pas. C'est cette comparaison qui permet, de façon intuitive, de déterminer si l'intervention a un effet causal sur la réponse. Si la réponse diffère sous les deux conditions, nous affirmons que l'intervention a un effet causal (adverse ou préventif) sur la réponse. Dans le cas contraire, nous affirmons que l'intervention n'a pas d'effet causal sur la réponse.

L'inférence causale est un champ de recherche multidisciplinaire, prenant notamment racine dans la statistique, l'épidémiologie, l'économétrie et les sciences sociales. C'est une discipline en plein essor qui s'est rapidement développée au cours des trois dernières décennies. Bien que la recherche en inférence causale ait atteint une certaine maturité, on trouve encore des défis importants dans ce domaine. Par exemple, un des principaux objectifs statistiques de l'inférence causale appliquée aux données observationnelles est de construire des estimateurs sans biais de l'effet de la variable de traitement sur la réponse d'intérêt. Ces estimateurs requièrent notamment l'identification des variables confondantes pour éviter le biais de confusion. En effet, ces variables confondantes, qui sont des causes communes de la variable de traitement et de la variable de réponse, induisent une association fallacieuse que l'on doit éliminer. Très populaire en inférence causale, le score de propension est un outil utilisé pour contrôler le biais de confusion. Si l'inclusion des confondants et des purs prédicteurs de la réponse dans le modèle du score de propension permet de construire un estimateur sans biais et précis de l'effet total de l'exposition sur la réponse, l'inclusion des purs prédicteurs de l'exposition ou des variables non reliées à la réponse et à l'exposition nuit à la qualité de l'estimation (Brookhart *et al.*, 2006; De Luna *et al.*, 2011; Shortreed et Ertefaie, 2017).

La sélection des variables pour l'inférence causale, que nous désignons par le terme général « sélection de confondants », est un problème difficile en pratique. Il existe une littérature de plus en plus volumineuse sur la sélection des confondants. Certains travaux supposent que la vraie structure des données est connue, tandis que d'autres sont basés sur des algorithmes d'apprentissage statistique et ne nécessitent aucune connaissance a priori de ladite structure de données. Les méthodes de sélection des confondants basées sur les algorithmes d'apprentissage statistique sont très utiles en pratique, car la vraie nature des variables n'est habituellement pas connue. Par ailleurs, une caractéristique commune à la majorité des algorithmes basés sur les données est que ceux-ci ont été développés pour des problèmes en petite dimension, c'est-à-dire lorsque la taille de l'échantillon est grande comparée au nombre de covariables. En effet, relativement peu de travaux sur la sélection des confondants ciblent spécifiquement des problèmes

en grande dimension, c'est-à-dire lorsque le nombre de variables augmente avec la taille de l'échantillon. La sélection de confondants en grande dimension est un défi de plus en plus rencontré en inférence causale, dû à la numérisation croissante des domaines d'application de l'inférence causale. Le problème de la sélection des confondants en grande dimension est particulièrement complexe, car on doit tenir compte de la sparsité et de la colinéarité simultanément.

Un autre grand défi, qui a moins retenu l'attention jusqu'à maintenant, est la sélection des confondants en analyse de la médiation causale. L'analyse de médiation causale est une technique statistique qui aide à comprendre comment le traitement affecte la réponse à travers une variable intermédiaire appelée médiateur. Plus particulièrement, un objectif principal de l'analyse de médiation est la décomposition de l'effet total du traitement en un effet direct et indirect. Dans un contexte observationnel, c'est-à-dire lorsque le traitement et le médiateur ne sont pas sous le contrôle de l'investigateur, cette décomposition est généralement erronée s'il y un contrôle inadéquat sur les variables confondantes. En analyse de médiation causale, l'ensemble des variables confondantes à cibler est élargi et inclut additionnellement les confondants de la relation médiateur-réponse, ce qui complexifie davantage l'identification et la sélection des variables.

Dans cette thèse, nous nous intéressons à l'estimation de l'effet causal d'un traitement sur une réponse sur la base de données observationnelles. Nous abordons plus particulièrement le problème des algorithmes de sélection des confondants basés sur les données, avec une emphase sur les problèmes en grande dimension et pour l'analyse de la médiation causale. Le premier chapitre effectue une introduction aux concepts fondamentaux de l'inférence causale. Le reste de la thèse (Chapitres 2-4) est divisé en trois grandes parties, qui sont décrites dans ce qui suit.

Le premier projet s'intitule « generalized outcome-adaptive lasso (GOAL) ». Ce projet a été accepté dans le journal *Biometrics* sous la forme d'une note « réaction du lecteur » de l'article de Shortreed et Ertefaie (2017, *Biometrics* 73(4), 111-1122). Dans cet article important, Shortreed et Ertefaie (2017) ont proposé l'approche « outcome-adaptive lasso

(OAL) » pour la sélection des variables dans un contexte d'inférence causale. Le OAL est une technique très astucieuse basée sur le lasso adaptatif qui permet de sélectionner les bonnes variables à inclure dans le modèle de score de propension et exclure celles qui sont nuisibles. Malgré la bonne performance de OAL en petite dimension et avec des prédicteurs non corrélés, sa performance se dégrade en grande dimension et pour les prédicteurs corrélés. C'est dans ce contexte que nous avons proposé GOAL, afin de généraliser et améliorer OAL. Le GOAL est basé sur la pénalisation dite « elastic net adaptatif » au lieu du lasso adaptatif, menant à une performance améliorée sur des données de grande dimension ou corrélées.

Le deuxième projet s'intitule « outcome-adaptive regularization for causal mediation analysis (MOAR) ». Les approches OAL et GOAL ont été conçues spécifiquement pour l'estimation de l'effet total du traitement sur la réponse. Comme nous l'avons précédemment mentionné, les chercheurs s'intéressent à l'estimation de l'effet total, mais aussi souvent à l'estimation de l'effet direct et indirect du traitement à travers un médiateur spécifique. Il est alors pertinent de se demander comment les connaissances et algorithmes développés pour la sélection des variables pour l'estimation de l'effet total se transposent dans un contexte de médiation. C'est dans ce cadre de travail que nous nous sommes intéressés à développer une nouvelle technique basée sur une application à deux étapes des approches OAL de Shortreed et Ertefaie (2017) et GOAL de Baldé et al. (2022), qui permet de sélectionner les covariables les plus appropriées pour l'analyse de la médiation causale. La méthode MOAR a été illustrée avec des données réelles provenant de l'étude Harvard School of Public Health College Alcohol. Ce travail est en voie d'être soumis sous la forme d'un article original dans le journal *Statistics in Medicine*.

Le troisième projet s'intitule « graphical lasso networks in the context of causal mediation analysis ». Le lasso graphique est une approche de régularisation très populaire et rapide qui a été développée par Friedman, Hastie et Tibshirani (2008) pour estimer l'inverse de la matrice de variance-covariance à partir de données normales multivariées. Cette matrice inverse est intéressante puisqu'elle encode les associations conditionnelles

entre les variables, ce qui peut être utile pour comprendre les relations entre celles-ci. Dans le cadre de ce projet, nous avons souhaité étudier la valeur du lasso graphique en analyse de médiation causale et proposer un ajustement afin de mieux tenir compte des particularités du contexte d'analyse ciblé. Un objectif concret, en voie de production, est de proposer une interface graphique simple qui permet aux utilisateurs de posséder un outil de visualisation des relations directes et de médiation probables entre variables.

Nous concluons cette thèse par une discussion afin de mettre en perspective les travaux de recherche présentés dans les chapitres précédents.

CHAPITRE I

INFÉRENCE CAUSALE

Dans ce chapitre, nous présentons les concepts de l'inférence causale qui sont essentiels pour aborder cette thèse. Plus concrètement, nous faisons une revue de littérature des approches contrefactuelles, des approches graphiques et de l'analyse de la médiation causale.

1.1 Approches contrefactuelles

Neyman (1923) introduit les approches contrefactuelles pour étudier des données d'expériences randomisées. Dans une expérience randomisée, l'attribution dans un groupe (traitement ou contrôle) se fait de façon aléatoire (par tirage au sort). Cette attribution aléatoire permet d'avoir une comparabilité optimale des deux groupes pour réduire les erreurs ou les biais statistiques. C'est pour cette raison que les données provenant d'une étude randomisée sont préférables en inférence causale. Cependant, il est parfois très coûteux ou impossible de conduire une étude randomisée. Dans ce cas une étude observationnelle peut être envisagée. Il est donc important de développer des approches équivalentes pour les données d'observation. Rubin (1974) généralise les approches contrefactuelles pour étudier des données observationnelles. Le matériel présenté dans cette section est tiré ou inspiré des travaux de Hernán et Robins (2020), Holland (1986), Lunceford et Davidian (2004) et Pearl (2009).

1.1.1 Définitions et notations mathématiques

Soient A la variable de traitement (exposition) observée ($A = 1$ si traité, $A = 0$ si non traité), X le vecteur des covariables mesurées avant la réception du traitement et Y la variable réponse observée. On associe à chaque individu de la population un vecteur aléatoire (Y_0, Y_1), où Y_1 (respectivement Y_0) est la réponse qu'on pourrait observer si l'individu avait reçu le traitement (respectivement si l'individu n'avait pas reçu le traitement). Les variables Y_1 et Y_0 sont appelées réponses contrefactuelles ou réponses potentielles, car on ne peut les observer simultanément. La réponse observée Y peut être écrite comme

$$Y = Y_1 A + Y_0(1 - A). \tag{1.1}$$

L'Équation (1.1) est appelée hypothèse de cohérence « consistency ». Nous donnerons plus de détails sur cette hypothèse dans la suite.

L'objectif est d'estimer l'effet du traitement (A) sur la réponse (Y) en utilisant les données observées (X, A, Y). Dans un premier temps, il est tout à fait raisonnable de s'intéresser à estimer l'effet causal individuel (ECI) du traitement (A) sur la réponse (Y).

Définition 1.1.1. *L'effet causal individuel est défini par*

$$ECI = Y_1 - Y_0.$$

L'effet causal individuel compare les deux réponses contrefactuelles Y_0 et Y_1. Si $ECI \neq 0$, alors on dit que le traitement (A) a un effet causal sur la réponse (Y) pour l'individu i. Par contre si $ECI = 0$, alors le traitement (A) n'a pas d'effet causal sur la réponse (Y) pour cet individu. L'effet causal individuel n'est pas identifiable car une des deux réponses contrefactuelles est toujours manquante. Une alternative pourrait alors être de considérer une population et d'estimer l'effet causal moyen (ECM).

Définition 1.1.2. *L'effet causal moyen de la population est défini par*

$$ECM = E(Y_1) - E(Y_0).$$

La quantité $E(Y_1)$ est la réponse contrefactuelle moyenne qu'on pourrait observer si tous les individus de la population avaient reçu le traitement ($A = 1$), tandis que $E(Y_0)$ représente la réponse contrefactuelle moyenne qu'on pourrait observer si tous les individus de la population n'avaient pas reçu le traitement ($A = 0$). Le ECM compare les deux réponses contrefactuelles moyennes $E(Y_1)$ et $E(Y_0)$. Si $ECM \neq 0$, alors on dit que A a un effet causal moyen sur Y pour la population, sinon ($ECM = 0$) on dit que A n'a pas un effet causal moyen sur Y. Ici aussi on ne peut pas identifier l'ECM car un individu ne peut pas être traité et non traité en même temps.

Les problèmes de données manquantes rencontrés dans l'estimation du ECI et ECM sont appelés le problème fondamental de l'inférence causale « Fundamental Problem of Causal Inference ». Le problème fondamental de l'inférence causale est que, en général, ECI et ECM ne peuvent pas être identifiés sans hypothèses supplémentaires, en raison de données manquantes. Dans ce sens, l'inférence causale peut être vue comme un problème de données manquantes.

Données randomisées

Le problème des données manquantes peut être résolu par une expérience randomisée. Dans une expérience randomisée, le traitement est attribué de façon aléatoire conduisant à une indépendance marginale ou conditionnelle entre les réponses contrefactuelles (Y_0, Y_1) et le traitement A. L'expérience aléatoire marginale garantit l'échangeabilité marginale $(Y_0, Y_1) \amalg A$, où $\amalg$ dénote l'indépendance statistique; c'est-à-dire, les réponses contrefactuelles (Y_0, Y_1) et le traitement A sont indépendants et donc les groupes traités et non traités sont échangeables sans condition. Une expérience randomisée conditionnelle combine deux (ou plusieurs) expériences randomisées marginales distinctes menées dans différents sous-ensembles de la population caractérisés par le vecteur X. Elle garantit uniquement une échangeabilité conditionnelle dans les niveaux de la variable X (c'est-à-dire $(Y_0, Y_1) \amalg A \mid X$). Dans la proposition qui suit nous montrons comment on peut

identifier l'ECM dans une expérience randomisée marginale. On sait que pour identifier l'ECM, il suffit d'identifier $E(Y_1)$ et $E(Y_0)$.

Proposition 1.1.1. *Dans une expérience aléatoire randomisée marginale on a :*

$$E(Y_1) = E(Y|A = 1) \quad et \quad E(Y_0) = E(Y|A = 0).$$

Démonstration.

$$\begin{aligned} E(Y|A=1) &= E(Y_1A + Y_0(1-A)|A=1) && \text{par l'Équation (1.1)} \\ &= E(Y_1|A=1) && \text{par la connaisance de } A=1 \\ &= E(Y_1) && \text{par la randomisation de } A\text{: } (Y_0, Y_1) \amalg A. \end{aligned}$$

De façon analogue on trouve $E(Y|A = 0) = E(Y_0)$.

□

Données observationnelles

Dans une étude randomisée, le traitement est attribué aléatoirement, mais dans une étude observationnelle le traitement peut être déterminé par de nombreux facteurs. Lorsque ces facteurs affectent le risque de développer la réponse, alors les effets de ces facteurs se confondent avec l'effet du traitement (Hernán et Robins, 2020). On parle alors de confusion pour la relation causale entre le traitement A et la réponse Y.

Définition 1.1.3. *La confusion pour la relation causale entre le traitement (A) et la réponse (Y) est une forme de manque d'échangeabilité entre le groupe traité et le groupe non traité.*

La confusion est souvent considérée comme la principale préoccupation des études observationnelles. L'inférence causale à partir d'une étude observationnelle est basée sur plusieurs hypothèses: positivité, cohérence, échangeabilité et l'hypothèse de stabilité de la valeur du traitement pour l'unité. Ces quatre hypothèses sont nécessaires pour proposer des estimateurs sans biais et convergents de l'ECM.

Hypothèses 1.1.1.

1. *La **positivité** peut être écrite sous la forme* $0 < P(A = 1|X = x) < 1$ *pour toutes les valeurs* x *possibles. Cela signifie que la probabilité de recevoir les deux niveaux de traitement conditionnellement à* X *est positive pour tous les individus.*

2. *La **cohérence** est définie comme* $Y = AY_1 + (1-A)Y_0$*. C'est-à-dire que la réponse observée pour un individu est égale à la réponse contrefactuelle dans le cadre de l'affectation de traitement que l'individu a réellement reçue.*

3. *L'**échangeabilité** (hypothèse d'ignorabilité forte) signifie que* X *inclut tous les confondants possibles :* $(Y_0, Y_1) \amalg A \mid X$*.*

4. *L'**hypothèse de stabilité de la valeur du traitement pour l'unité** « stable unit treatment value assumption » est définie en deux étapes: i) il y a une seule version de traitement* $(A = 1)$ *et contrôle* $(A = 0)$*; ii) il n'y pas d'interférence entre les individus. On dit qu'il y a interférence lorsque le traitement (exposition) reçu par un individu affecte les réponses contrefactuelles d'autres individus.*

Lorsque les hypothèses de 1 à 4 sont satisfaites, on peut conceptualiser l'étude observationnelle comme une expérience randomisée conditionnelle. Autrement dit, si les hypothèses 1 à 4 sont satisfaites, on peut identifier $E(Y_1)$ et $E(Y_0)$ à partir des données observées (X, A, Y). En effet,

Démonstration.

$$\begin{aligned}
E\{E(Y|A=1,X)\} &= \sum_x E(Y|A=1,X=x)P(X=x) \\
&= \sum_x E(Y_1A + Y_0(1-A)|A=1,X=x)P(X=x) \quad \text{par cohérence} \\
&= \sum_x E(Y_1|X=x,A=1)P(X=x) \quad \text{par la connaisance de } A=1 \\
&= \sum_x E(Y_1|X=x)P(X=x) \quad \text{par} \quad (Y_0,Y_1) \amalg A|X \\
&= E(Y_1).
\end{aligned}$$

□

De la même façon on trouve $E\{E(Y|A = 0, X)\} = E(Y_0)$. Alors l'$ECM$ peut être identifié si les hypothèses 1 à 4 sont satisfaites.

Il existe des méthodes basées sur le score de propension qui permettent d'estimer l'ECM. Parmi ces méthodes on peut mentionner les techniques d'appariement, de stratification, de pondération par probabilité inverse « Inverse Probability Weighting (IPW) » et l'estimation double robuste. On peut trouver une revue de littérature détaillée de ces dites méthodes dans Hernán et Robins (2020) et Lunceford et Davidian (2004). Dans le cadre de cette thèse nous allons rappeler seulement la méthode IPW (Lunceford et Davidian, 2004).

1.1.2 Score de propension

Lorsque la dimension du vecteur des covariables X est grande, il peut être difficile d'inclure toutes les covariables dans le modèle de regression et donc d'estimer avec précision les deux fonctions de régressions $E\{E(Y|A = 1, X)\}$ et $E\{E(Y|A = 0, X)\}$ (Hirano et Imbens, 2001). La méthodologie du score de propension « propensity score (PS) » est une approche très populaire qui a été introduite par Rosenbaum et Rubin (1983) pour résoudre ce problème (Hirano et Imbens, 2001).

Définition 1.1.4. *Le score de propension $e(X)$ est la probabilité qu'un individu soit traité conditionnellement à ses variables pré-traitement X (covariables):*

$$e(X) = P(A = 1|X). \tag{1.2}$$

Le score de propension a des propriétés très importantes (Rosenbaum et Rubin, 1983). Avant de citer ces propriétés, nous allons d'abord définir le score d'équilibre « balancing score ».

Définition 1.1.5. *Un score d'équilibre « balancing score », $b(X)$, est une fonction des covariables observées X telles que la distribution conditionnelle de X étant donné $b(X)$*

est la même pour les individus traitées (A = 1) et les individus non traités $(A = 0)$.

Propriétés du score de propension (Rosenbaum et Rubin, 1983)
(i) Le score de propension $e(X)$ est un score d'équilibre:

$$X \amalg A \mid e(X);$$

(ii) Si $(Y_0, Y_1) \amalg A \mid X$ alors

$$(Y_0, Y_1) \amalg A \mid e(X);$$

(iii) Le score de propension est un scalaire $(0 < e(X) < 1)$, il ne dépend pas de la dimension de X.

La propriété (i) signifie que conditionnellement à $e(X)$, la distribution de X est indépendante de A. La propriété (ii) signifie que si le traitement (A) est fortement ignorable conditionnellement aux covariables (X) alors le traitement (A) est fortement ignorable conditionnellement au score de propension. Autrement dit, si l'hypothèse 3.) (hypothèse d'ignorabilité forte) est satisfaite, le traitement est indépendant des réponses contrefactuelles pour les individus qui partagent le même score de propension. La propriété (iii) signifie que le score de propension est une probabilité comprise strictement entre 0 et 1 et est indépendante de la dimension du vecteur des covariables (X). En général, le score de propension est inconnu pour les données d'observation et doit être estimé. La régression logistique est souvent utilisée pour estimer le score de propension. Dans ce cas, on peut considérer un modèle logistique paramétré par α:

$$e(X) = \{1 + \exp(-X^T\alpha)\}^{-1}, \tag{1.3}$$

où α est un vecteur colonne de $\mathbb{R}^p$.

1.1.3 Pondération par probabilité inverse « Inverse Probability Weighting »

Le « Inverse Probability Weighting » (IPW) est une méthode basée sur le score de propension très utilisée en inférence causale. Elle crée une pseudo-population dans laquelle il n'y a pas de lien causal entre l'exposition (A) et les covariables (X), de sorte

que les moyennes pondérées reflètent les moyennes dans la vraie population. Les Figures 1.1 et 1.2 illustrent graphiquement la méthode IPW. Ces deux figures sont présentées pour donner un aspect visuel de la méthode IPW, mais une revue de littérature détaillée sur les approches graphiques est présentée dans la Section 1.2.

Avant la pondération IPW, dans la Figure 1.1, X est une cause commune de A et Y donc un confondant. Après la pondération, dans la Figure 1.2, X ne cause plus A, donc X n'est plus un confondant de A et Y.

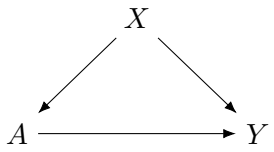


Figure 1.1 Diagramme causal avant la pondération (IPW)

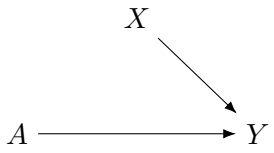


Figure 1.2 Diagramme causal après la pondération (IPW)

L'estimateur IPW de l'ECM est donné par

$$\widehat{ECM}_{IPW} = \frac{1}{n}\sum_{i=1}^{n}\frac{A_i Y_i}{\hat{e}(X_i)} - \frac{1}{n}\sum_{i=1}^{n}\frac{(1-A_i)Y_i}{(1-\hat{e}(X_i))}, \tag{1.4}$$

où $\hat{e}(X_i)$ est le score de propension estimé pour l'individu i et n le nombre d'individus. Sous certaines conditions, $\widehat{ECM}_{IPW}$ est un estimateur convergent de l'ECM (Lunceford et Davidian, 2004). En supposant que $\hat{e}(X)$ est un estimateur convergent de $e(X)$, il suffit de montrer que

$$E\left\{\frac{AY}{e(X)}\right\} = E(Y_1) \quad \text{et} \quad E\left\{\frac{(1-A)Y}{1-e(X)}\right\} = E(Y_0). \tag{1.5}$$

Démonstration.

Premièrement, les espérances à gauche dans (1.5) sont bien définies puisque par l'hypothèse de **positivité**, on a $e(X) \neq 0$ et $1 - e(X) \neq 0$ car $0 < e(X) < 1$. De plus,

$$\begin{aligned}
E\left\{\frac{AY}{e(X)}\right\} &= E\left[E\left\{\frac{AY}{e(X)}\middle| X\right\}\right] && \text{par le conditionnement sur } X \\
&= E\left[E\left\{\frac{AY_1}{e(X)}\middle| X\right\}\right] && \text{par l'hypothèse de } \textbf{cohérence} \\
&= E\left[E(Y_1|X)\frac{E(A|X)}{e(X)}\right] && \text{par l'}\textbf{échangeabilité}\text{: } (Y_0, Y_1) \amalg A|X \\
&= E\left[E(Y_1|X)\frac{e(X)}{e(X)}\right] && \text{car } E(A|X) = e(X) \\
&= E\left[E(Y_1|X)\right] \\
&= E(Y_1).
\end{aligned}$$

Par analogie, on trouve que $E\left\{\frac{(1-A)Y}{1-e(X)}\right\} = E(Y_0)$.

□

Dans cette présente section, nous nous sommes concentrés sur les hypothèses et sur les méthodes pour calculer l'effet du traitement sur la réponse d'intérêt. Lorsque la structure des données est simple, il n'est pas nécessaire d'expliciter le diagramme causal. Cependant, lorsque nous faisons face à des situations plus complexes, l'utilisation des graphiques causaux est utile, sinon nécessaire. Les approches graphiques permettent de clarifier les problèmes conceptuels en inférence causale (Hernán et Robins, 2020).

1.2 Approches graphiques

Le premier diagramme causal découle des travaux du biologiste et statisticien américain Wright (1921). L'utilisation des approches graphiques est largement répandue et a été fortement influencée par les travaux importants de Spirtes, Glymour et Scheines (2000) et Pearl (2009). Dans cette section, nous présentons les approches graphiques pour représenter les principaux concepts causaux. Ces concepts sont essentiels pour une meilleure compréhension du Chapitre 4, dans lequel nous développons une approche graphique basée sur le lasso pour l'identification des confondants dans un contexte de

l'analyse de la médiation causale. Sauf mention contraire, le contenu de cette section est inspiré des livres de Spirtes, Glymour et Scheines (2000), Pearl (2009) et Hernán et Robins (2020).

1.2.1 Définitions

La Figure 1.3 est tirée du livre de Hernán et Robins (2020). Ce graphe est composé de trois noeuds représentant trois variables données (A, Y, L) et trois flèches ($\rightarrow$). Dans la Figure 1.3, la présence de la flèche pointant de A vers Y indique qu'il existe un effet causal direct de A sur Y pour au moins un individu dans la population. A contrario, une absence de flèche de A vers Y signifie que A n'a aucun effet causal direct sur Y pour tous les individus de la population.

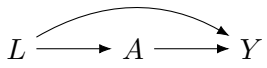


Figure 1.3 Graphe orienté acyclique causal.

Le type de diagramme présenté à la Figure 1.3 est un graphe orienté acyclique ou DAG « directed acyclic graph ». Ce graphe est orienté car toutes les arêtes (liens) entre variables possèdent une direction. De plus, il est acyclique car en partant de n'importe quel noeud (L, A ou Y), il est impossible de revenir au même noeud en suivant le sens des flèches. Un DAG est causal si les causes communes de toute paire de variables dans le graphique sont également dans le graphique (Hernán et Robins, 2020). La variable L dans la Figure 1.3 est une cause commune de A et Y.

Un *chemin* entre deux variables dans un DAG est un itinéraire qui relie les variables par des flèches de telle sorte que l'itinéraire ne visite aucune variable plus d'une fois (Hernán et Robins, 2020). Un *chemin porte arrière* de A à Y est un chemin entre A et Y où il y a une flèche qui pointe sur A. Dans la Figure 1.3, le chemin qui relie A et Y à travers la cause commune L est un exemple de chemin porte arrière ($A \leftarrow L \rightarrow Y$). Un *collisionneur* « collider » est un nœud (variable) sur un chemin dans lequel deux flèches

entrent en collision. Par exemple dans la Figure 1.3, la variable Y (effet commun des variables L et A) est un collisionneur dans le chemin $L \rightarrow Y \leftarrow A$. Un chemin est fermé (bloqué) s'il contient un collisionneur. Par exemple, le chemin $L \rightarrow Y \leftarrow A$ est fermé par la présence du collisionneur Y sur ce chemin. Un chemin qui ne contient pas de collisionneur est dit *ouvert*. Le chemin $A \leftarrow L \rightarrow Y$ est ouvert car il ne contient aucun collisionneur.

1.2.2 Association marginale

Nous utilisons des DAGs simples afin d'illustrer les situations où deux variables A et Y sont marginalement associées:

1. L'une cause l'autre. La Figure 1.4 illustre un DAG impliquant une association marginale entre les variables A et Y. La variable A cause la variable Y alors A et Y sont associées car la causalité implique l'association, mais pas l'inverse.

Figure 1.4 Association marginale basée sur un effet causal direct.

2. L'une cause l'autre à travers une autre. Dans la Figure 1.5, la variable L est une variable intermédiaire dans le chemin de A à Y ($A \rightarrow L \rightarrow Y$). La variable A cause indirectement la variable Y (à travers L) et on a donc une association entre A et Y.

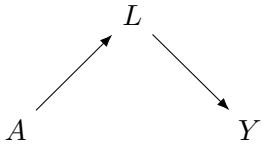


Figure 1.5 Association marginale basée sur un effet causal indirect.

3. Elles ont une cause commune. Dans la Figure 1.6, la variable L est la cause commune de A et Y.

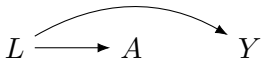


Figure 1.6 Association marginale basée sur une cause commune.

Dans tous les autres cas, les variables A et Y sont marginalement indépendantes. La Figure 1.7 montre un exemple où les variables A et Y sont marginalement indépendantes dû à la présence du collisionneur L dans le chemin entre A et Y ($A \rightarrow L \leftarrow Y$).

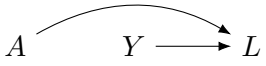


Figure 1.7 Indépendance marginale.

1.2.3 Association conditionnelle

Dans cette sous-section, une boîte autour d'une variable indique un conditionnement sur cette variable.

Nous utilisons des DAGs simples afin d'illustrer les situations où deux variables A et Y sont associées conditionnellement à L:

1. Elles ont un effet commun et l'analyse conditionne sur celui-ci ou se limite à certains niveaux de cet effet commun. Dans la Figure 1.8, la variable L est un collisionneur.

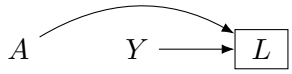


Figure 1.8 Association conditionnelle.

2. Elles ont un effet commun qui a un descendant et l'analyse conditionne sur celui-ci ou se limite à certains niveaux de ce descendant. Dans la Figure 1.9, la variable L est un collisionneur et la variable C est un descendant du collisionneur L.

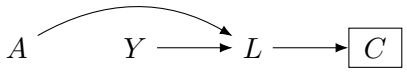


Figure 1.9 Association conditionnelle avec descendant.

Il faut également remarquer que le conditionnement sur la variable L ferme les chemins (présentement ouverts) $A \rightarrow L \rightarrow Y$ dans la Figure 1.5 et $A \leftarrow L \rightarrow Y$ dans la Figure 1.6, respectivement. Ceci implique une indépendance conditionnelle de A et Y.

1.3 Analyse de la médiation causale

Soit (A, M, Y) le triplet de variables de traitement (exposition, intervention ou prévention), de médiateur et de réponse d'intérêt. L'analyse de la médiation causale cherche à expliquer le mécanisme par lequel la variable traitement (A) affecte la variable de réponse d'intérêt (Y).

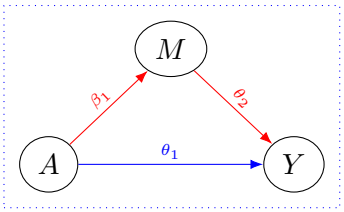


Figure 1.10 Graphe de médiation causale.

Le graphe de médiation causale présenté en Figure 1.10 fait l'hypothèse selon laquelle la variable indépendante (A) cause une variable intermédiaire (M), qui à son tour cause la variable dépendante (Y). La variable intermédiaire M est appelée médiateur.

L'analyse de médiation cherche à décomposer l'effet total en un effet indirect et un effet direct. Dans la Figure 1.10, l'effet direct est traduit par le chemin $A \rightarrow Y$, tandis que l'effet indirect est traduit par le chemin $A \rightarrow M \rightarrow Y$. Dans un article important, Baron et Kenny (1986) proposent une approche paramétrique pour estimer les effets direct et indirect (Valeri et VanderWeele, 2013). Pour le cas d'un médiateur et d'une réponse

continus, nous considérons les modèles de régression suivants:

$$E[Y|A=a, X=x] = \phi_0 + \phi_1 a + \phi_2' x \tag{1.6}$$

$$E[Y|A=a, M=m, X=x] = \theta_0 + \theta_1 a + \theta_2 m + \theta_4' x \tag{1.7}$$

$$E[M|A=a, X=x] = \beta_0 + \beta_1 a + \beta_2' x. \tag{1.8}$$

Les coefficients de régression β_1, θ_1 et θ_2 sont représentés à la Figure 1.10.

1.3.1 Approches traditionelles

Baron et Kenny (1986) propose une approche en quatre étapes pour déterminer si un médiateur potentiel médie effectivement la relation entre le traitement et la réponse:

1. Il existe une association entre A et M, c'est-à-dire que dans le modèle (1.8) on a $\beta_1 \neq 0$.

2. Il existe une association entre M et Y, ce qui se traduit par $\theta_2 \neq 0$ dans le modèle (1.7).

3. Il existe une association entre A et Y (sans contrôle sur M), c'est-à-dire que, dans le modèle (1.6), on a $\phi_1 \neq 0$.

4. Après contrôle sur M, il n'existe plus d'association entre A et Y. Ceci signifie que $\theta_1 = 0$ dans le modèle (1.7).

Il faut noter qu'en pratique, les coefficients $(\beta_1, \theta_2, \phi_1, \theta_1)$ sont estimés et les critères sont vérifiés sur la base de la significativité statistique de ces coefficients.

Un consensus a été établi et seuls les deux premiers critères ont été acceptés comme critères valides (Valeri et VanderWeele, 2013). Pour identifier M comme médiateur potentiel, les critères 3 et 4 ne sont pas nécessaires. En effet, le critère 3 a été rejeté car l'association entre A et Y pourrait être nulle lorsque les effets direct et indirect sont de signe opposé (médiation incohérente). Le critère 4 a été rejeté pour la simple raison que

la médiation peut être complète ou partielle (Valeri et VanderWeele, 2013). Alors que seul l'effet indirect est non nul dans une médiation complète, pour la médiation partielle les deux effets (direct et indirect) sont non nuls. Lorsque la médiation est partielle, une association de A et Y peut toujours être présente après le contrôle sur M, mais la force de la relation est typiquement réduite.

Il existe deux approches traditionnelles principales pour l'analyse de la médiation, à savoir la méthode du produit et la méthode par la différence (Valeri et VanderWeele, 2013). La méthode du produit est généralement référée comme étant l'approche de Baron et Kenny (1986) (Valeri et VanderWeele, 2013). La méthode du produit définit l'effet direct par θ_1 et l'effet indirect par $\beta_1\theta_2$. Autrement dit, l'effet indirect est le chemin par lequel le traitement affecte le médiateur de β_1 et le médiateur à son tour affecte la réponse (conditionnellement à l'exposition) de θ_2, $(A \to M \to Y)$. Alors que la méthode par différence a le même effet direct que la méthode du produit (θ_1), elle définit l'effet indirect par $\phi_1 - \theta_1$, où ϕ_1 représente l'effet total de l'exposition sur la réponse.

Les méthodes traditionnelles ont plusieurs limites, parmi lesquelles on peut citer les deux limitations les plus importantes. Premièrement, les approches traditionnelles manquent de généralité, car elles ne permettent pas l'inclusion de termes d'interaction traitement-médiateur (Valeri et VanderWeele, 2013). Lorsqu'une interaction traitement-médiateur est présente et que l'approche traditionnelle est utilisée, les estimations d'effet obtenues ne sont plus interprétables comme des effets direct et indirect. Deuxièmement, les approches traditionnelles ignorent souvent l'hypothèse d'absence de confondants médiateur-réponse non mesurés (Valeri et VanderWeele, 2013; VanderWeele, 2015). Cette hypothèse doit être satisfaite tant pour les données randomisées (pour l'exposition) qu'observationnelles. En effet, la randomisation de l'exposition, en comparaison à la randomisation du médiateur, ne permet pas de mitiger la confusion entre le médiateur et la réponse. Il est alors important de développer des approches alternatives qui prennent pleinement en considération les aspects causaux de l'analyse de médiation.

1.3.2 Approches causales

Soit Y_a la réponse contrefactuelle pour chaque individu lors de l'intervention pour fixer le traitement A à a. Soit M_a le médiateur contrefactuel pour chaque individu lorsqu'on intervient pour fixer le traitement A à a. Soit Y_{am} la réponse contrefactuelle de Y pour chaque individu lorsqu'on intervient pour fixer le traitement A à a et le médiateur M à m. Soit $Y_{aM_{a^*}}$, la réponse contrefactuelle emboîtée, la réponse de l'individu si le traitement est fixé à $A = a$ et que le médiateur M est fixé au niveau qu'il devrait être si le traitement était fixé à $A = a^*$.

Dans ce qui suit, nous allons définir les effets contrôlé et naturels.

Définition 1.3.1. *L'effet direct contrôlé « controlled direct effect (CDE) » est défini par*

$$CDE = E[Y_{1m}] - E[Y_{0m}].$$

Le CDE exprime à quel point la réponse changerait en moyenne si le médiateur était contrôlé au niveau m uniformément dans la population mais que le traitement passait du niveau $A = 0$ au niveau $A = 1$.

Définition 1.3.2. *L'effet naturel direct « natural direct effect (NDE) » est défini par*

$$NDE = E[Y_{1M_0}] - E[Y_{0M_0}].$$

Le NDE est la différence en moyenne des réponses contrefactuelles entre le groupe de traitement $A = 1$ et le groupe de contrôle $A = 0$ tout en contrôlant le médiateur au niveau de M_0, le niveau que le médiateur serait en absence de traitement.

Définition 1.3.3. *L'effet naturel indirect « natural indirect effect (NIE) » est défini par*

$$NIE = E[Y_{1M_1}] - E[Y_{1M_0}].$$

Le NIE est la différence en moyenne des réponses contrefactuelles entre le niveau du médiateur qui serait obtenu dans la condition de traitement (M_1) et dans la condition de contrôle (M_0), en maintenant le traitement au niveau $A = 1$.

Proposition 1.3.1. *L'effet total « total effect (TE) » se décompose en NDE et NIE:*

$$TE = NIE + NDE.$$

Démonstration.

$$\begin{aligned} TE &= E[Y_1] - E[Y_0] \\ &= E[Y_{1M_1}] - E[Y_{0M_0}] \quad \text{par l'hypothèse de composition (VanderWeele, 2015)} \\ &= E[Y_{1M_1} - Y_{1M_0}] + E[Y_{1M_0} - Y_{0M_0}] \\ &= NIE + NDE. \end{aligned}$$

□

Hypothèses 1.3.1. *Les hypothèses suivantes sont nécessaires pour l'identification du NDE et NIE:*

H_1 *: Il n'existe pas de confondants traitement-réponse non mesurés;*

H_2 *: Il n'existe pas de confondants médiateur-réponse non mesurés;*

H_3 *: Il n'existe pas de confondants traitement-médiateur non mesurés;*

H_4 *: Il n'existe pas de confondants médiateur-réponse impacté par l'exposition.*

Identification des effets CDE, NDE et NIE

Dans cette sous-section nous présentons l'identifcation des effets CDE, NDE et NIE par la formule de médiation en supposant un médiateur M discret (VanderWeele, 2015; Pearl, 2009; Imai *et al.*, 2010a). Notons que dans ce qui suit, les effets sont exprimés conditionnellement aux covariables X.

En supposant que les hypothèses H_1-H_2 sont satisfaites l'effet CDE peut être identifié par:

$$CDE = E(Y|A = 1, M = m, X = x) - E(Y|A = 0, M = m, X = x).$$

Sous les hypothèses H_1-H_4 les effets NDE et NIE peuvent être identifiés par

$$\begin{aligned} NDE = &\sum_m [E(Y|A = 1, M = m, X = x) - E(Y|A = 0, M = m, X = x)] \\ &\times P(M = m|A = 0, X = x); \\ NIE = &\sum_m [P(M = m|A = 1, X = x) - P(M = m|A = 0, X = x)] \\ &\times E(Y|A = 1, M = m, X = x). \end{aligned}$$

Nous référons le lecteur intéressé pour plus de détails aux travaux de VanderWeele (2015), de Imai et al. (2010a) et de Pearl (2009).

Plusieurs approches ont été développées pour estimer les effets contrôlé, naturel direct et naturel indirect (Vansteelandt, 2009; VanderWeele et Vansteelandt, 2009; Tingley *et al.*, 2014; VanderWeele, 2015; Steen *et al.*, 2017; Valeri et VanderWeele, 2013). Plus spécifiquement, Tingley et al. (2014) proposent le package `R` « Mediation » pour estimer les effets (NDE et NIE) par simulation. Plus tard, Steen, Loeys, Moerkerke et Vansteelandt (2017) développent un autre package `R`, « medflex », pour estimer les effets par pseudo-population. D'autres auteurs ont proposé une estimation des effets (NDE et NIE) par régression (VanderWeele, 2015; Valeri et VanderWeele, 2013). Dans le cadre de cette thèse, nous avons utilisé l'approche séquentielle de Vansteelandt (2009) « Sequential g-estimation » pour l'effet direct moyen contrôlé « controlled direct effect (CDE) ». En supposant qu'il n'existe pas de confondants affectés par le traitement, l'identification de

l'effet CDE nécessite seulement les hypothèses H_1 et H_2 (Vansteelandt, 2009).

Régression pour l'analyse de la médiation causale

Soient $a = 1$ (traité) et $a^* = 0$ (non traité) les niveaux du traitement A. Considérons les modèles de régression suivants:

$$E[M|A = a, X = x] = \beta_0 + \beta_1 a + \beta_2' x \tag{1.9}$$

$$E[Y|A = a, M = m, X = x] = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4' x. \tag{1.10}$$

Si les hypothèses H_1-H_4 sont satisfaites, les effets CDE, NDE et NIE peuvent être obtenus par:

$$CDE = (\theta_1 + \theta_3 m)(a - a^*)$$

$$NDE = \{\theta_1 + \theta_3(\beta_0 + \beta_1 a^* + \beta_2' x)\}(a - a^*)$$

$$NIE = (\theta_2\beta_1 + \theta_3\beta_1 a)(a - a^*).$$

Les estimateurs de CDE, NDE et NIE peuvent alors être définis par (VanderWeele et Vansteelandt, 2009; VanderWeele, 2015; Valeri et VanderWeele, 2013):

$$\widehat{CDE} = (\hat{\theta}_1 + \hat{\theta}_3 m)(a - a^*)$$

$$\widehat{NDE} = \{\hat{\theta}_1 + \hat{\theta}_3(\hat{\beta}_0 + \hat{\beta}_1 a^* + \hat{\beta}_2' x)\}(a - a^*)$$

$$\widehat{NIE} = (\hat{\theta}_2\hat{\beta}_1 + \hat{\theta}_3\hat{\beta}_1 a)(a - a^*),$$

où $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3$ et $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ sont des estimateurs de $\theta_1, \theta_2, \theta_3$ et $\beta_0, \beta_1, \beta_2$, respectivement.

Si le modèle linéaire (1.10) n'inclut pas de terme d'interaction traitement-médiateur ($\theta_3 = 0$) alors on voit que $CDE = NDE = \theta_1(a - a^*)$, c'est-à-dire les effets directs naturel et contrôlé sont égaux. Les approches traditionnelles et les approches causales coïncident dans certains contextes et divergent dans d'autres (Valeri et VanderWeele, 2013). Par exemple, pour les modèles linéaires et log-linéaires, les approches traditionnelles et les approches causales coïncident lorsqu'il n'y a pas d'interaction traitement-médiateur (Valeri et VanderWeele, 2013). Pour les modèles logistiques, les approches traditionnelles

et les approches contrefactuelles coïncident lorsqu'il n'y a pas d'interaction exposition-médiateur et lorsque la réponse est relativement rare (VanderWeele et Vansteelandt, 2009; VanderWeele et Vansteelandt, 2010; Valeri et VanderWeele, 2013). S'il y a une interaction traitement-médiateur, alors l'approche contrefactuelle doit être utilisée (Valeri et VanderWeele, 2013). Pour plus de détails sur les méthodes par régression nous renvoyons le lecteur à Valeri et VanderWeele (2013), à VanderWeele (2015) et à Samoilenko et Lefebvre (2021).

Estimation g-séquentielle « Sequential g-estimation » (Vansteelandt, 2009)

Dans cette sous-section, nous rappelons brièvement la méthode d'estimation séquentielle de Vansteelandt (2009). L'approche d'estimation g-séquentielle se concentre principalement sur une estimation de l'effet direct contrôlé moyen (CDE) en supprimant l'effet du médiateur (M) sur la réponse (Y). Cette approche séquentielle de Vansteelandt (2009) peut être utilisée en présence de confondants intermédiaires, qui sont des confondants médiateur-réponse affectés par le traitement. Soit Z les confondants intermédiaires. Dans la suite, nous présentons les trois étapes séquentielles de l'estimation g-séquentielle:

1. Premièrement, estimer l'effet de M sur Y à l'aide de la régression suivante:

$$(\hat{\beta}_A^{ols}, \hat{\beta}_M^{ols}, \hat{\beta}^{ols}) = \arg \min_{(\beta_A, \beta_M, \beta)} \|Y - \beta_A A - \beta_M M - \mathbf{X}\beta\|_2^2$$

   ou (en présence des confondants intermédiaires)

$$(\hat{\beta}_A^{ols}, \hat{\beta}_Z^{ols}, \hat{\beta}_M^{ols}, \hat{\beta}^{ols}) = \arg \min_{(\beta_A, \beta_M, \beta)} \|Y - \beta_A A - \beta_Z Z - \beta_M M - \mathbf{X}\beta\|_2^2 ;$$

2. Ensuite, l'effet de M est supprimé de Y pour définir une nouvelle variable de réponse Y_{res} exempté de l'effet de M:

$$Y_{res} = Y - \hat{\beta}_M^{ols} M;$$

3. Enfin, le CDE (encodé par γ_A) est estimé par la régression :

$$Y_{res} = \gamma_A A + \gamma X + \nu, \quad \nu \sim N(0, \sigma_\nu^2 \mathbf{I}).$$

CHAPITRE II

PREMIER ARTICLE: READER REACTION TO " OUTCOME-ADAPTIVE LASSO: VARIABLE SELECTION FOR CAUSAL INFERENCE " BY SHORTREED AND ERTEFAIE (2017)

Ismaila Baldé, Yi Archer Yang, Geneviève Lefebvre

Les méthodes de sélection de variables pour la prédiction ont largement été étudiées dans la littérature statistique, tandis qu'il y a comparativement peu de travaux sur la sélection de variables pour l'inférence causale. Dans cette thèse nous nous concentrons plus spécifiquement sur la sélection de variables pour l'inférence causale basée sur des données d'observation. En particulier, nous nous concentrons sur l'estimation sans biais de l'effet du traitement sur la réponse en présence de confondants, tout en maintenant l'efficacité statistique.

Ce Chapitre 2 est formé d'un article dont je suis le premier auteur. Cet article est publié au journal *Biometrics*. Pour cet article, j'ai collaboré avec ma directrice de thèse, Geneviève Lefebvre, professeure au Département de mathématiques de l'UQAM et Yi Archer Yang, professeur au Département de mathématiques et statistique de l'Université McGill. J'ai développé l'idée originale, effectué la revue de la littérature, réalisé les analyses statistiques et rédigé le code informatique et le manuscrit.

2.1 Préliminaires: Méthodes de sélection de variables régularisées

La sélection de variables est extrêmement importante pour la découverte de connaissances avec des problèmes de grande dimension (Fan et Li, 2006). On parle de problèmes de grande dimension (ou problèmes de haute dimension) « high-dimensional problems » lorsque le nombre de variables (p) augmente avec la taille de l'échantillon (n). Pour simplifier la présentation, on utilisera seulement l'expression « grande dimension ». En raison de la numérisation croissante des domaines de la biomédecine, des sciences sociales et de l'économie, la sélection de variables avec des données de grande dimension devient un sujet d'intérêt croissant. De plus, la sélection des variables dans un tel contexte est très difficile, puisqu'il faut tenir compte simultanément de la sparsité et de la corrélation entre les covariables.

Considérons le modèle de régression linéaire classique suivant avec un vecteur réponse observé Y et une matrice de covariables $\mathbf{X} = (X_1, \ldots, X_p)$:

$$Y = \mathbf{X}\beta + \epsilon,$$

où $\epsilon \sim N(0, \sigma^2\mathbf{I}_n)$ et β est le vecteur de coefficients inconnus de la régression. Sans perte de généralité, nous supposons que les données sont centrées et que l'ordonnée à l'origine n'est pas incluse dans le modèle de régression.

2.1.1 Estimateur des moindres carrés ordinaires « Ordinary least squares (OLS) »

L'estimateur des moindres carrés ordinaires est défini par

$$\hat{\beta}^{ols} = \arg\min_{\beta} \left(\|Y - \mathbf{X}\beta\|_2^2 \right). \tag{2.1}$$

L'approche OLS estime généralement des coefficients non nuls de β. Lorsque que la dimension p du vecteur de covariables est grande, on s'attend à ce que certaines composantes de β soient exactement zéro. Autrement dit, le vrai modèle a une représentation sparse. Dans ce contexte, des approches de sélection de variables sont nécessaires pour estimer le modèle sparse. Au cours des trois dernières décennies, plusieurs méthodes

de régularisation ont été proposées pour effectuer simultanément la sélection des variables et l'estimation des coefficients correspondants. Le lasso (Tibshirani, 1996) est l'une des méthodes les plus populaires en raison de ses bonnes propriétés statistiques et de ses avantages computationnels. L'utilisation du lasso a beaucoup été influencée par l'apparition de l'algorithme « least angle regression (LARS) » (Efron *et al.*, 2004) et les algorithmes de descente de coordonnées « coordinate descent algorithms » (Tibshirani, 2011). L'algorithme LARS offre une façon simple et efficace de résoudre le lasso (Tibshirani, 2011). Les algorithmes de descente de coordonnées sont extrêmement simples et rapides, et exploitent la sparsité supposée du modèle de manière très avantageuse (Tibshirani, 2011).

Dans la suite de cette section, nous passons brièvement en revue le lasso (Tibshirani, 1996) et ses variantes qui sont utiles pour comprendre les méthodes et algorithmes développés dans cette thèse. Pour plus de détails, nous renvoyons le lecteur intéressé aux travaux de Tibshirani (1996), Zou et Hastie (2005), Zou (2006), Zou et Zhang (2009), Ghosh (2011) et Fan et Li (2001).

2.1.2 Le lasso (Tibshirani, 1996)

Tibshirani (1996) a défini l'estimateur lasso comme suit

$$\hat{\beta} = \arg\min_{\beta} \left(\|Y - \mathbf{X}\beta\|_2^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right),$$

où λ est un paramètre de régularisation non négatif ($\lambda \geq 0$). Le lasso est une méthode utilisant une pénalité basée sur la norme L_1. Lorsque le paramètre de régularisation (λ) est choisi de manière appropriée, l'algorithme lasso réduit les coefficients de certaines covariables à zéro. Une grande valeur de λ mène à une plus grande pénalisation, tandis qu'une petite valeur de λ mène à une petite réduction des coefficients des variables. Malgré les bonnes performances de la méthode lasso, elle ne satisfait pas une propriété asymptotique très importante, appelée propriété de l'oracle « oracle property » (Fan et Li, 2001; Zou, 2006; Zou et Zhang, 2009; Ghosh, 2011). En effet, Fan et Li (2001) ont

d'abord souligné que le lasso présente asymptotiquement un biais non négligeable pour l'estimation des coefficients non nuls avant de conjecturer que le lasso ne satisfaisait pas la propriété d'oracle à cause du problème de biais (Fan et Li, 2001; Zou et Zhang, 2009). Cette dite conjecture a été prouvée par Zou (2006) (Zou, 2006; Zou et Zhang, 2009). Zou (2006) utilise le langage de Fan et Li (2001) et appelle δ une procédure d'oracle si $\hat{\beta}(\delta)$ (asymptotiquement) a les propriétés d'oracle suivantes :

(i) Identifie le bon modèle de sous-ensemble, $\{j : \hat{\beta}_j \neq 0\} = \mathcal{A}$;

(ii) A le taux d'estimation optimal, $\sqrt{n}\left(\hat{\beta}(\delta)_{\mathcal{A}} - \beta_{\mathcal{A}}\right) \xrightarrow{d} N(0, \mathbf{\Sigma})$,

où $\mathcal{A} = \{j : \beta_j \neq 0\}$ et $\mathbf{\Sigma}$ est la matrice de covariance connaissant le vrai modèle de sous-ensemble. Il est établi qu'une bonne procédure devrait avoir ces propriétés d'oracle (Fan et Li, 2001; Fan et Peng, 2004; Zou, 2006).

2.1.3 Le lasso adaptatif « adaptive lasso » (Zou, 2006)

Zou (2006) a défini l'estimateur lasso adaptatif comme suit

$$\hat{\beta} = \arg\min_{\beta} \left(\|Y - \mathbf{X}\beta\|_2^2 + \lambda \sum_{j=1}^{p} \tilde{w}_j |\beta_j| \right),$$

où $\tilde{w}_j = |\tilde{\beta}_j|^{-\gamma}$ tel que $\gamma > 0$ et $\tilde{\beta}$ est un estimateur $\sqrt{n}$-convergent de β. Zou (2006) a prouvé qu'avec un λ choisi de manière appropriée, le lasso adaptatif satisfait la propriété de l'oracle. En d'autres termes, l'estimateur lasso adaptatif satisfait les conditions de convergence dans la sélection des variables (identifie le modèle de sous-ensemble de variables correct) et la normalité asymptotique (Zou, 2006; Zou et Zhang, 2009; Ghosh, 2011). Cependant, le lasso adaptatif hérite de l'instabilité du lasso pour les données en grande dimension (Zou et Zhang, 2009). Les problèmes de colinéarité sont courants dans les contextes d'analyse statistique en grande dimension (Zou et Zhang, 2009). En effet, Fan et Lv (2008) montrent que même lorsque les covariables sont indépendantes, la corrélation échantillonnale maximale entre covariables peut être importante si la dimension est grande. C'est pourquoi on dit souvent que le problème de grande dimension

induit souvent le problème de colinéarité (Zou et Zhang, 2009). Zou et Hastie (2005) ont développé l'approche elastic net pour améliorer le lasso pour les données en grande dimension et les données corrélées.

2.1.4 L'elastic net (Zou et Hastie, 2005)

La méthode de l'elastic net effectue simultanément une régularisation et une sélection de variables, avec une pénalité qui est une combinaison convexe des pénalités lasso et ridge (Zou et Hastie, 2005) (ridge utilise une pénalité basée sur la norme L_2). L'avantage de l'elastic net est que le lasso résout le problème de sparsité du modèle et que le ridge s'attaque au problème de la collinéarité. Soient λ_1 et λ_2 deux paramètres de régularisation non négatifs. Zou et Hastie (2005) définissent l'estimateur elastic net par

$$\hat{\beta} = \arg\min_{\beta} \left(\|Y - \mathbf{X}\beta\|_2^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=1}^{p} \beta_j^2 \right).$$

D'une part le lasso adaptatif a un problème avec la collinéarité et d'autre par l'elastic net ne satisfait pas la propriété de l'oracle (Zou et Zhang, 2009). Motivés par le développement d'une approche appropriée pour faire face aux problèmes de grande dimension, Zou et Zhang (2009) proposent l'estimateur elastic net adaptatif « adaptive elastic net » pour améliorer le lasso adaptatif et l'elastic net.

2.1.5 L'elastic net adaptatif « adaptive elastic net » (Zou et Zhang, 2009; Ghosh, 2011)

La méthode de l'elastic net adaptatif a été spécifiquement introduite pour analyser les données de grande dimension (Zou et Zhang, 2009). Il s'agit d'une combinaison du lasso adaptatif et de l'elastic net, héritant de bonnes propriétés de chacune des deux méthodes: la propriété d'oracle du premier et la capacité de surmonter le problème de colinéarité de la seconde (Zou et Zhang, 2009; Ghosh, 2011). Ghosh (2011) a défini l'estimateur

elastic net adaptatif comme suit

$$\hat{\beta} = \arg\min_{\beta} \left(\|Y - \mathbf{X}\beta\|_2^2 + \lambda_1 \sum_{j=1}^{p} \tilde{w}_j |\beta_j| + \lambda_2 \sum_{j=1}^{p} \beta_j^2 \right),$$

où $\tilde{w}_j = \left|\tilde{\beta}_j\right|^{-\gamma}$ tel que $\tilde{\beta}$ est un estimateur $\sqrt{n}$-convergent de β et $\gamma > 0$. Par exemple, $\hat{\beta}^{ols}$ défini en (2.1) est un estimateur $\sqrt{n}$-convergent de β et donc peut être utilisé pour construire les poids $\tilde{w}$.

Comme pour l'elastic net (Zou et Hastie, 2005), Ghosh (2011) a montré l'équivalence entre l'elastic net adaptatif (Zou et Zhang, 2009) et un problème lasso adaptatif standard dans un espace augmenté (voir Web Appendix A dans Baldé et al. (2022) pour la preuve) :

$$\hat{\beta} = \arg\min_{\beta} \left(\|Y^* - \mathbf{X}^*\beta\|_2^2 + \lambda_1 \sum_{j=1}^{p} \tilde{w}_j |\beta_j| \right),$$

où $\mathbf{X}^* = \begin{pmatrix} \mathbf{X} \\ \sqrt{\lambda_2}\mathbf{I}_p \end{pmatrix}$, $Y^* = \begin{pmatrix} Y \\ 0_p \end{pmatrix}$.

2.1.6 La sélection des variables en inférence causale

Le problème de sélection des variables est aussi souvent rencontré en inférence causale sur des données d'observation. Alors que certains chercheurs considèrent le problème de sélection des variables sur la base d'une bonne connaissance de la vraie nature des données observées (le vrai graphe causal des données), d'autres proposent des algorithmes d'apprentissage qui ne supposent aucune connaissance de la vraie structure des données. En pratique, la vraie nature des données est généralement inconnue et donc doit être estimée.

La méthode du score de propension (PS) est souvent utilisée pour contrôler le biais de confusion dans les études observationnelles. Dans le Chapitre 1, nous avons expliqué que le PS peut être estimé par une régression logistique (pour un traitement binaire). Donc il est possible de penser à utiliser des méthodes de classification par apprentis-

sage automatique, le lasso ou ses variantes pour estimer le PS. Dans ce contexte, il faut rappeler que le lasso et ses variantes tels que définis précédemment ont pour objectif d'estimer des relations parcimonieuses pour le traitement étant donné les covariables. Toutefois, l'objectif de la sélection de variables pour le score de propension est d'estimer un score d'équilibre parcimonieux pour tenir compte du biais de confusion, tout en maintenant l'efficacité statistique. C'est pourquoi il est sous-optimal d'utiliser directement le lasso ou ses variantes pour sélectionner les variables du score de propension. Il est ainsi particulièrement important de développer des méthodes de sélection de variables pour l'inférence causale.

Récemment, Shortreed et Ertefaie (2017) ont proposé le outcome-adaptive lasso (OAL). Le OAL est une version modifiée du lasso adaptatif de Zou (2006) en pénalisant moins les variables associées à la réponse (pure prédicteurs de la réponse et les confondants) pour encourager leur sélection (Ye *et al.*, 2021). Baldé et al. (2022) généralisent OAL (GOAL) pour faire face aux problèmes de sélection des confondants en grande dimension à travers une modification de l'elastic net adaptatif de Zou et Zhang (2009) et Ghosh (2011), soit le outcome-adaptive elastic net, pour améliorer OAL de Shortreed et Ertefaie (2017) dans les problèmes de grande dimension. D'autres auteurs ont récemment et indépendamment repris l'idée de l'elastic net adaptatif basé sur la réponse pour l'inférence causale à partir de données d'observation en grande dimension (Islam et Noor-E-Alam, 2021).

Reader Reaction to " Outcome-adaptive lasso: Variable selection for causal inference " by Shortreed and Ertefaie (2017)

Ismaila Baldé, Yi Yang, Geneviève Lefebvre

Abstract: Shortreed and Ertefaie (2017, Biometrics 73(4), 1111 − 1122) introduced a clever propensity score variable selection approach for estimating average causal effects, namely the outcome adaptive lasso (OAL). OAL aims to select desirable covariates, confounders and predictors of outcome, to build an unbiased and statistically efficient propensity score estimator. Due to its design, a potential limitation of OAL is how it handles the collinearity problem, which is often encountered in high-dimensional data. As seen in Shortreed and Ertefaie (2017), OAL's performance degraded with increased correlation between covariates. In this note, we propose the generalized outcome adaptive lasso (GOAL) that combines the strengths of the adaptively weighted L_1 penalty and the elastic net to better handle the selection of correlated covariates. Two different versions of GOAL, which differ in their procedure (algorithm), are proposed. We compared OAL and GOAL in simulation scenarios that mimic those examined by Shortreed and Ertefaie (2017). While all approaches performed equivalently with independent covariates, we found that both GOAL versions were more performant than OAL in low and high dimensions with correlated covariates.

Keywords: Adaptive elastic net; Causal inference; High dimensional data; Propensity score; Variable selection.

2.2 Introduction

In a very interesting paper, Shortreed and Ertefaie (2017, Biometrics **73(4)**, 1111–1122) introduced the outcome adaptive lasso (OAL) approach for variable selection in the causal inference framework. OAL was designed to target confounders and predictors of outcome, while excluding spurious covariates and covariates only associated with

exposure. As explained therein, variables selected that way aim to yield an unbiased and efficient propensity score (PS) estimator. Shortreed and Ertefaie (2017) theoretically and empirically demonstrated that OAL is able to select all true confounders and predictors of outcome, and exclude the rest of covariates. The performance of the algorithm was examined in situations wherein the number of predictors was small or large relative to the number of observations. Indeed the authors suggested that OAL is also adequate to be used in high-dimensional problems (p increasing with n), which are common in causal inference.

In high dimension, an ideal variable selection approach should enjoy the oracle property and deal with the collinearity problem (Zou and Zhang, 2009) which typically plagues such settings. OAL is based on the adaptive lasso method (Zou, 2006) and features the oracle property, but its ability to properly treat correlated predictors is questionable. As seen in simulation scenarios presented in Shortreed and Ertefaie (2017), OAL was observed increasingly biased and variable as the correlation between predictors increased (see Web Appendices on: https://onlinelibrary.wiley.com/doi/10.1111/biom.12679). Similar degraded performance is known for lasso, where unstable solution paths are obtained when predictors are highly correlated (Zou and Hastie, 2005).

It has previously been shown that elastic net can overcome the collinearity problem exhibited by lasso (Zou and Hastie, 2005). Moreover, it is possible to transform the elastic net problem into an equivalent lasso problem by augmenting the data (Zou and Hastie, 2005). This last idea was transposed to adaptive elastic net, which permits variable selection consistency and encourages grouping effect that is, either selection or omission of correlated variables together (Ghosh, 2011). Building on these works, we propose an outcome adaptive elastic net approach to improve on OAL, which we name the generalized outcome adaptive lasso (GOAL). Our idea is to start from an outcome adaptive elastic net problem that can be transformed into Shortreed and Ertefaie's outcome adaptive lasso representation by augmenting the data. Two different versions of our approach, namely Naive GOAL and GOAL with PIRLS, were explored.

The first version solves the GOAL problem naively using the `R` function *lqa*. Developed by Ulbricht (2010) to fit penalized generalized linear models, *lqa* was used by Shortreed and Ertefaie (2017) to solve the OAL problem. To implement the proposed Naive GOAL estimator, only straightforward data manipulations are required before calling the `R` code provided by Shortreed and Ertefaie (2017) for OAL. The second version solves the GOAL problem via a penalized iteratively re-weighted least squares (PIRLS) procedure. GOAL with PIRLS is based on a modified *lqa* function (referred herein as *mlqa*), which modifies the working response and weights of the Newton-Raphson update in the *lqa* function. To implement GOAL with the PIRLS estimator, we substitute the *lqa* function by *mlqa* in the `R` function.

In our work, which picks up simulation scenarios presented in Shortreed and Ertefaie (2017), both versions of GOAL were observed to perform similarly to OAL when predictors were uncorrelated. However GOAL was seen to offer better performance than OAL when correlation between predictors was present.

Our note is structured as follows. Section 2.3 contains an overview of the methods. We present the GOAL approach in Section 2.3.1 and provide two simple algorithms (versions) for its implementation in Section 2.3.2. We describe the simulation study in Section 2.4 and corresponding results are presented in Section 2.5. We conclude with a discussion in Section 2.6.

2.3 Methods

We introduce the GOAL approach along with the data augmentation step which underlies both versions of our GOAL estimator. In our presentation, we adopt Shortreed and Ertefaie (2017)'s notation to describe variables and models. More precisely, we let $(\mathbf{X}, A, Y)$ denote the triplet of design matrix, treatment and response, respectively, where $\mathbf{X} = (X_1, X_2, \ldots, X_p)$.

2.3.1 Generalized outcome adaptive lasso (GOAL)

We briefly recall Shortreed and Ertefaie (2017) method (OAL) before introducing the proposed approach (GOAL). We assume the following PS model parametrized by α:

$$logit\left\{P(A=1|\mathbf{X})\right\} = \sum_{j=1}^{p} \alpha_j X_j.$$

The objective of OAL is to estimate the following PS model:

$$logit\left\{\pi(X,\hat{\alpha})\right\} = \sum_{j\in\mathcal{C}} \hat{\alpha}_j X_j + \sum_{j\in\mathcal{P}} \hat{\alpha}_j X_j,$$

where $\mathcal{C}$ and $\mathcal{P}$ denote indices of confounders (covariates related to both outcome and treatment) and pure predictors of outcome, respectively.
Let $\ell_n(\alpha; A, \mathbf{X}) = \sum_{i=1}^{n} \left\{-a_i(x_i^T\alpha) + \log\left(1 + e^{x_i^T\alpha}\right)\right\}$ be the negative log-likelihood. OAL is an adaptive lasso penalty for logistic regression; it is defined as

$$\hat{\alpha}(OAL) = \arg\min_{\alpha} \left[\ell_n(\alpha; A, \mathbf{X}) + \lambda_n \sum_{j=1}^{p} \hat{w}_j |\alpha_j|\right], \tag{2.2}$$

where $\hat{w}_j = \left|\hat{\beta}_j^{ols}\right|^{-\gamma}$ such that $\gamma > 1$ and $(\hat{\beta}_A^{ols}, \hat{\beta}^{ols}) = \arg\min_{(\beta_A,\beta)} \|Y - \beta_A A - \mathbf{X}\beta\|_2^2$. Now let us define $\mathcal{A} = \mathcal{C} \cup \mathcal{P} = \{1, 2, \ldots, p_0\}$, the indices of desirable covariates to include in the estimated PS. Let $\mathcal{A}^c = \mathcal{I} \cup \mathcal{S} = \{p_0 + 1, p_0 + 2, \ldots, p_0 + (p - p_0)\}$ be the indices of of covariates to exclude, where $\mathcal{I}$ and $\mathcal{S}$ are pure predictors of treatment and spurious covariates, respectively. Thus, we can define the Fisher Information matrix as

$$\mathbf{F}(\alpha^*) = \begin{pmatrix} \mathbf{F}_{11} & \mathbf{F}_{12} \\ \mathbf{F}_{21} & \mathbf{F}_{22} \end{pmatrix},$$

where $\mathbf{F}_{11}$ is the Fisher Information matrix (of size $p_0 \times p_0$) for the parsimonious PS. In the following theorem, Shortreed and Ertefaie (2017) show that with a proper choice of λ_n and γ, the OAL approach enjoys the oracle property.

Theorem 2.3.1. *(Shortreed et Ertefaie, 2017)*

Suppose $\lambda_n/\sqrt{n} \to 0$ *and* $\lambda_n n^{\gamma/2-1} \to \infty$, *for* $\gamma > 1$. *Then the outcome-adaptive lasso estimator* $\hat{\alpha}(OAL)$ *satisfies the following:*

1. *Consistency in variable selection:* $\lim_n P\{\hat{\alpha}_j(OAL) = 0 | j \in \mathcal{I} \cup \mathcal{S}\} = 1$.

2. *Asymptotic normality:* $\sqrt{n}\{\hat{\alpha}(OAL) - \alpha^*_{\mathcal{A}}\} \to_d N(0, \mathbf{F}_{11}^{-1})$.

As seen in Theorem 2.3.1, the OAL method forces coefficients corresponding to pure predictors of treatment and spurious variables to zero. Moreover, the OAL estimator is asymptotically unbiased. The interested reader can find the proof of Theorem 2.3.1 in the Supplementary Materials of the paper of Shortreed and Ertefaie (2017).
Shortreed and Ertefaie (2017) used the `R` function *lqa* to solve problem (2.2). Moreover, they proposed to minimize a weighted absolute mean difference (wAMD) between the treated and untreated groups to select the tuning parameter λ_n in the set

$$S_{\lambda_n} = \{n^{-10}, n^{-5}, n^{-2}, n^{-1}, n^{-0.75}, n^{-0.5}, n^{-0.25}, n^{0.25}, n^{0.49}\},$$

that is $\hat{\lambda}_n = \arg\min_{\lambda_n \in S_{\lambda_n}} wAMD(\lambda_n; \mathbf{X}, A)$, where

$$wAMD(\lambda_n; \mathbf{X}, A) = \sum_{j=1}^{p} \left|\hat{\beta}_j^{ols}\right| \left| \frac{\sum_{i=1}^n \hat{\tau}_i^{\lambda_n} X_{ij} A_i}{\sum_{i=1}^n \hat{\tau}_i^{\lambda_n} A_i} - \frac{\sum_{i=1}^n \hat{\tau}_i^{\lambda_n} X_{ij}(1 - A_i)}{\sum_{i=1}^n \hat{\tau}_i^{\lambda_n}(1 - A_i)} \right| \tag{2.3}$$

and $\hat{\tau}_i^{\lambda_n}$ is the inverse probability of treatment weight for individual i constructed using the PS model fitted from Equation (2.2).

Building upon the adaptive elastic net estimator (see Web Appendix A for a review), we define the GOAL estimator through the following optimization problem

$$\hat{\alpha}(GOAL) = \arg\min_{\alpha} \left[\ell_n(\alpha; A, \mathbf{X}) + \lambda_1 \sum_{j=1}^{p} \hat{w}_j |\alpha_j| + \lambda_2 \sum_{j=1}^{p} \alpha_j^2 \right], \tag{2.4}$$

where $\hat{w}_j$ is defined as in (2.2).

In the sequel, we propose two ways to solve problem (2.4) and obtain GOAL estimates.

Naive GOAL

Lemma 1 in Zou and Hastie (2005) was initially proposed for linear models to reexpress the elastic net problem into a lasso penalty. Algamal and Lee (2015a) directly

applied this lemma for logistic regression to transform elastic net into a lasso problem on augmented data. Building on these works as well as on Ghosh (2011), Naive GOAL adopts an augmented adaptive lasso representation for logistic regression, as described below. Given the original design matrix and treatment data $(\mathbf{X}, A)$ and fixed (λ_1, λ_2), we create an augmented data set $(\mathbf{X}^*, A^*)$: $\mathbf{X}^* = \begin{pmatrix} \mathbf{X} \\ \sqrt{\lambda_2}\mathbf{I}_p \end{pmatrix}$, $A^* = \begin{pmatrix} A \\ 0_p \end{pmatrix}$, where $\mathbf{X}^* = (X_1^*, X_2^*, \ldots, X_p^*)$. Following Zou and Hastie (2005) and Algamal and Lee (2015a), we reexpress the GOAL estimator (2.4) as an OAL problem on the data $(\mathbf{X}^*, A^*)$:

$$\hat{\alpha}_N(GOAL) = \arg\min_{\alpha} \left[\ell_{n^*}(\alpha; A^*, \mathbf{X}^*) + \lambda_1 \sum_{j=1}^{p} \hat{w}_j |\alpha_j| \right], \tag{2.5}$$

where $n^* = n + p$.

The solution to (2.5) can be obtained using the *lqa* function without any further modification.

GOAL with PIRLS

Tibshirani (1996) solved the lasso problem for logistic regression by applying the original lasso algorithm for linear regresion at each step of the penalized iteratively re-weighted least squares (PIRLS) method. That is, the L_1-penalized logistic regression is viewed as a lasso-weighted least squares (lasso-WLS) problem at each iteration of the PIRLS algorithm. To solve the OAL problem, Shortreed and Ertefaie (2017) relied upon the *lqa* (Ulbricht, 2010) function to fit the penalized logistic likelihood using the PIRLS technique. Let $\tilde{\alpha}$ be the current estimate of α in the PIRLS and ℓ_Q be the quadratic approximation of ℓ_n. The Newton-Raphson update solution of OAL is obtained as

$$\hat{\alpha}_{PIRLS}(OAL) = \arg\min_{\alpha} \left[\ell_Q(\alpha; A, \mathbf{X}, Z, \mathbf{T}) + \lambda_1 \sum_{j=1}^{p} \hat{w}_j |\alpha_j| \right],$$

where $\ell_Q(\alpha; A, \mathbf{X}, Z, \mathbf{T}) = \frac{1}{2} \sum_{i=1}^{n} t_i \left(z_i - x_i^T \alpha\right)^2$, $\quad \tilde{p}(x_i) = \frac{1}{1+\exp(-x_i^T \tilde{\alpha})}$,

$t_i = \tilde{p}(x_i)[1 - \tilde{p}(x_i)]$, $\quad z_i = x_i^T \tilde{\alpha} + \frac{a_i - \tilde{p}(x_i)}{\tilde{p}(x_i)(1-\tilde{p}(x_i))}$, $\quad Z = (z_1, \ldots, z_n)^T$, $\quad \mathbf{T} = (t_1, \ldots, t_n)$.

Similarly to OAL, the Newton-Raphson update solution of GOAL is obtained as

$$\hat{\alpha}_{PIRLS}(GOAL) = \arg\min_{\alpha} \left[\ell_Q(\alpha; A, \mathbf{X}, Z, \mathbf{T}) + \lambda_1 \sum_{j=1}^{p} \hat{w}_j |\alpha_j| + \lambda_2 \sum_{j=1}^{p} \alpha_j^2 \right] \tag{2.6}$$

$$= \arg\min_{\alpha} \left[\ell_{Q^*}(\alpha; A^*, \mathbf{X}^*, Z^*, \mathbf{T}^*) + \lambda_1 \sum_{j=1}^{p} \hat{w}_j |\alpha_j| \right] \tag{2.7}$$

where $\mathbf{X}^* = \begin{pmatrix} \mathbf{X} \\ \sqrt{\lambda_2}\mathbf{I}_p \end{pmatrix}$, $A^* = \begin{pmatrix} A \\ 0_p \end{pmatrix}$, $Z^* = \begin{pmatrix} Z \\ 0_p \end{pmatrix}$ and $\mathbf{T}^* = \begin{pmatrix} \mathbf{T} & 0_{n\times p} \\ 0^T_{n\times p} & \mathbf{I_p} \end{pmatrix}$.

We prove the equality between Equations (2.6) and (2.7) in Web Appendix B.

Recall that both GOAL estimators are modified versions of the adaptive elastic net (Ghosh, 2011; Zou and Zhang, 2009). Each estimator is reformulated as an OAL estimator in an augmented space, which is a modified version of adaptive lasso (Zou, 2006). Ghosh (2011) stated that the adaptive lasso formulation from adaptive elastic net is still a convex problem and enjoys the desirable optimization properties of adaptive lasso. Prior to this paper, Ghosh (2007) proved that the adaptive lasso in augmented space satisfies the oracle property with an additional assumption, namely $\lambda_2/n \to 0$. Additional discussions on oracle property for adaptive elastic net are available in Ghosh (2007), Ghosh (2011), Zou (2006), Zou and Zhang (2009) and Li et al. (2010). The following proposition combines results from Shortreed and Ertefaie (2017) and Ghosh (2007) to postulate the oracle property of GOAL.

Conjecture 2.3.1. *Suppose* $\lambda_1/\sqrt{n} \to 0$, $\lambda_1 n^{\gamma/2-1} \to \infty$, *for* $\gamma > 1$ *and* $\lambda_2/n \to 0$; *then the generalized outcome-adaptive lasso estimator* $\hat{\alpha}(GOAL)$ *satisfies the following:*

1. *Consistency in variable selection:* $\lim_n P\{\hat{\alpha}_j(GOAL) = 0 | j \in \mathcal{I} \cup \mathcal{S}\} = 1$;
2. *Asymptotic normality:* $\sqrt{n}\{\hat{\alpha}(GOAL) - \alpha^*_{\mathcal{A}}\} \to_d N(0, \mathbf{F}^{-1}_{11})$.

We conjecture the oracle property can be obtained by combining the proof in Zou (2006), Ghosh (2007) and Shortreed and Ertefaie (2017).

2.3.2 Implementation of GOAL

Selection of tuning parameters is a fundamental aspect of penalized model fitting. Since we defined the GOAL problem by using two tuning parameters (λ_1, λ_2), our proposal is to balance the exposure groups on a two-dimensional surface. Following Zou and Hastie (2005), we first consider relatively small values of λ_2:

$$S_{\lambda_2} = \{0, 10^{-2}, 10^{-1.5}, 10^{-1}, 10^{-0.75}, 10^{-0.5}, 10^{-0.25}, 10^{0}, 10^{0.25}, 10^{0.5}, 10^{1}\}.$$

Then for each fixed $\lambda_2 \in S_{\lambda_2}$, the GOAL algorithms solve the OAL problem with augmented data defined as a function of the λ_2 value. In this step, we recall that λ_1 is selected on the basis of the wAMD defined in Equation (2.3). The chosen (λ_1, λ_2) is the one maximizing the balance between exposure groups (that is, corresponding to the smallest wAMD). In the sequel, we summarize the steps to implement both versions of GOAL in Algorithms 1 and 2, respectively.

Algorithm 1 Naive GOAL

1: **Input**: Design matrix and treatment data $(\mathbf{X}, A)$

2: For each fixed λ_2 define: $\mathbf{X}^* = \begin{pmatrix} \mathbf{X} \\ \sqrt{\lambda_2}\mathbf{I}_p \end{pmatrix}$ and $A^* = \begin{pmatrix} A \\ 0_p \end{pmatrix}$

3: Call OAL algorithm with augmented data $(\mathbf{X}^*, A^*)$ to solve

$$\hat{\alpha}_N^*(\text{naive adaptive elastic net}) = \arg\min_{\alpha} \left[\ell_{n^*}(\alpha; \mathbf{X}^*, A^*) + \lambda_1 \sum_{j=1}^{p} \hat{w}_j |\alpha_j| \right]$$

4: Compute $\hat{\alpha}_N(\text{adaptive elastic net}) = (1 + \lambda_2)\hat{\alpha}_N^*(\text{naive adaptive elastic net})$

5: **Output**: $\hat{\alpha}_N(\text{adaptive elastic net})$

As seen above, while our first GOAL version (Algorithm 1) applies naively to a logistic model a data augmentation step that was originally developed for a linear model, our second GOAL version (Algorithm 2) performs the data augmentation within each iteration of the PIRLS. Also note that in steps 4 in Algorithm 1 and 7 in Algorithm 2,

Algorithm 2 GOAL with PIRLS

1: **Input**: Design matrix and treatment data $(\mathbf{X}, A)$

2: For each fixed λ_2 define: $\mathbf{X}^* = \begin{pmatrix} \mathbf{X} \\ \sqrt{\lambda_2}\mathbf{I}_p \end{pmatrix}$ and $A^* = \begin{pmatrix} A \\ 0_p \end{pmatrix}$

3: Initialize $\tilde{\alpha}$ to 0

4: Compute $\tilde{p}(x_i) = \frac{1}{1+\exp(-x_i^T\tilde{\alpha})}$, $t_i = \tilde{p}(x_i)[1-\tilde{p}(x_i)]$, $z_i = x_i^T\tilde{\alpha} + \frac{a_i - \tilde{p}(x_i)}{\tilde{p}(x_i)(1-\tilde{p}(x_i))}$, $i = 1, 2, \ldots, n$

5: Set $Z^* = \begin{pmatrix} Z \\ 0_p \end{pmatrix}$ and $\mathbf{T}^* = \begin{pmatrix} \mathbf{T} & 0_{n\times p} \\ 0_{n\times p}^T & \mathbf{I_p} \end{pmatrix}$, where $Z = (z_1, \ldots, z_n)^T, \mathbf{T} = (t_1, \ldots, t_n)$

6: Call OAL algorithm with augmented data $(\mathbf{X}^*, A^*)$ to solve

$$\hat{\alpha}_I^*(\text{naive adaptive elastic net}) = \arg\min_{\alpha} \left[\ell_{Q^*}(\alpha; \mathbf{X}^*, A^*, Z^*, \mathbf{T}^*) + \lambda_1 \sum_{j=1}^{p} \hat{w}_j |\alpha_j| \right]$$

7: Compute $\hat{\alpha}_I(\text{adaptive elastic net}) = (1+\lambda_2)\hat{\alpha}_I^*(\text{naive adaptive elastic net})$

8: Update $\tilde{\alpha} = \hat{\alpha}_I(\text{adaptive elastic net})$

9: Repeat $4-8$ until convergence of $\tilde{\alpha}$

10: Set $\tilde{\alpha}_I(\text{adaptive elastic net}) = \tilde{\alpha}$

11: **Output**: $\tilde{\alpha}_I(\text{adaptive elastic net})$

we compute the adaptive elastic net as discussed in Zou and Hastie (2005) and Ghosh (2011).

2.4 Simulation Study

The simulation study was designed to investigate the performance of GOAL, as compared to OAL, in higher and lower dimensional settings. We followed Shortreed and Ertefaie (2017) simulation setup to generate the data $(\mathbf{X}, A, Y)$. They simulated $X_i = (X_{i1}, X_{i2}, \ldots, X_{ip})_{1\leq i\leq n}$ from a multivariate standard Gaussian distribution with pairwise correlation ρ; binary treatment A from a Bernoulli distribution with $logit\{P(A_i = 1)\} = \sum_{j=1}^{p} \alpha_j X_{ij}$ and continuous outcome as $Y_i = \beta_A A_i + \sum_{j=1}^{p} \beta_j X_{ij} + \epsilon_i$ where $\epsilon_i \sim N(0,1)$ and $\beta_A = 0$ or 2. We varied both the sample size (n) and the number of

covariates (p). To evaluate GOAL's performance, we examined all (n, p) combinations of the original paper of Shortreed and Ertefaie (2017), that is: $n = 200$ with $p = 100$ and $n = 500$ with $p = 200$ for the high-dimensional settings, and $n = 200, 500, 1000$ with fixed $p = 20$ for the low-dimensional settings.

In our simulations, we considered the same four scenarios as in Shortreed and Ertefaie (2017). These scenarios are defined as follows, where $\beta \in \mathbb{R}^p$ are the regression coefficients in the outcome model and $\alpha \in \mathbb{R}^p$ are the regression coefficients in the treatment model:

(a) Scenario 1 sets $\beta = (0.6, 0.6, 0.6, 0.6, 0, 0, 0, \ldots, 0)$ and $\alpha = (1, 1, 0, 0, 1, 1, 0, \ldots, 0)$;

(b) Scenario 2 sets $\beta = (0.6, 0.6, 0.6, 0.6, 0, 0, 0, \ldots, 0)$ and $\alpha = (0.4, 0.4, 0, 0, 1, 1, 0, \ldots, 0)$;

(c) Scenario 3 sets $\beta = (0.2, 0.2, 0.6, 0.6, 0, 0, 0, \ldots, 0)$ and $\alpha = (1, 1, 0, 0, 1, 1, 0, \ldots, 0)$;

(d) Scenario 4 sets $\beta = (0.6, 0.6, 0.6, 0.6, 0, 0, 0, \ldots, 0)$ and $\alpha = (1, 1, 0, 0, 1.8, 1.8, 0, \ldots, 0)$.
In each scenario, the first two covariates are confounders, the third and fourth covariates are outcome predictors (unrelated to treatment), the fifth and sixth covariates are exposure predictors (unrelated to outcome) and the rest are spurious covariates (i.e. $p-6$ spurious covariates). Four different correlations ($\rho = 0, 0.2, 0.5, 0.75$) between covariates were investigated, where the first three values were considered by Shortreed and Ertefaie (2017). We refer the interested reader to the original paper (Shortreed and Ertefaie, 2017: Section 4.1, Section 6 and Web appendices) for more details on the simulation. For each scenario, we obtained estimates for the average treatment effect (ATE) using the IPTW estimator (Lunceford and Davidian, 2004) with the PS model fitted using either OAL or GOAL. We compared OAL and GOAL approaches based on the bias, standard error (SE) and mean squared error (MSE) of resulting IPTW estimators for the ATE. For variable selection, we used the proportion of times each predictor was selected for inclusion in the PS model (tolerance was 10^{-8}; the variable is selected when its coefficient is greater than 10^{-8}) under 1000 simulations.

2.5 Results

Figure 2.1 and Web Table 2.1 (first two sections) present results associated with Scenarios 1-2 in the high-dimensional settings ($p/n = 100/200, 200/500$), displaying the bias, SE and MSE of OAL and GOAL estimators for the ATE under a grid of increasing values for ρ (0, 0.2, 0.5, 0.75) when the true ATE is 0. Due to space constraints, the corresponding results for Scenarios 3-4 are presented in Web Figure 2.2 and Web Table 2.1 (last two sections). In addition, box plots of ATE estimates for OAL and GOAL are presented in Web Figures 2.3-2.6 by scenario and p/n ratios, separately for each ρ value. All the results for the low-dimensional settings ($n = 200, 500, 1000$ and $p = 20$) when the true ATE is 0 are presented in Web Appendix C (Web Table 2.2 and Web Figures 2.7-2.10). Selected complementary results for wAMD and variable selection in the high- and low-dimensional settings are also found in Web Appendix C. Since similar results were obtained when the true ATE was 2, we omit their presentation for all scenarios. In the sequel, we refer to Naive GOAL as GOALn and GOAL with PIRLS as GOALi.

In the high-dimensional settings ($p/n = 100/200, 200/500$), all three estimators (OAL, GOALn and GOALi) performed similarly when $\rho = 0$ (refer to Figure 2.1 (Scenarios 1-2), Web Figure 2.2 (Scenarios 3-4) and Web Table 2.1). When $\rho = 0.5$ or 0.75, GOAL was found systematically less variable than OAL, and either GOALi or both GOALi and GOALn exhibited less bias than OAL. GOALi's bias was small or relatively small for all ρ values. The difference between GOAL and OAL estimators was the largest when $\rho = 0.75$. Notably, the MSE of OAL was found at least twice the MSE of GOAL under this correlation value.

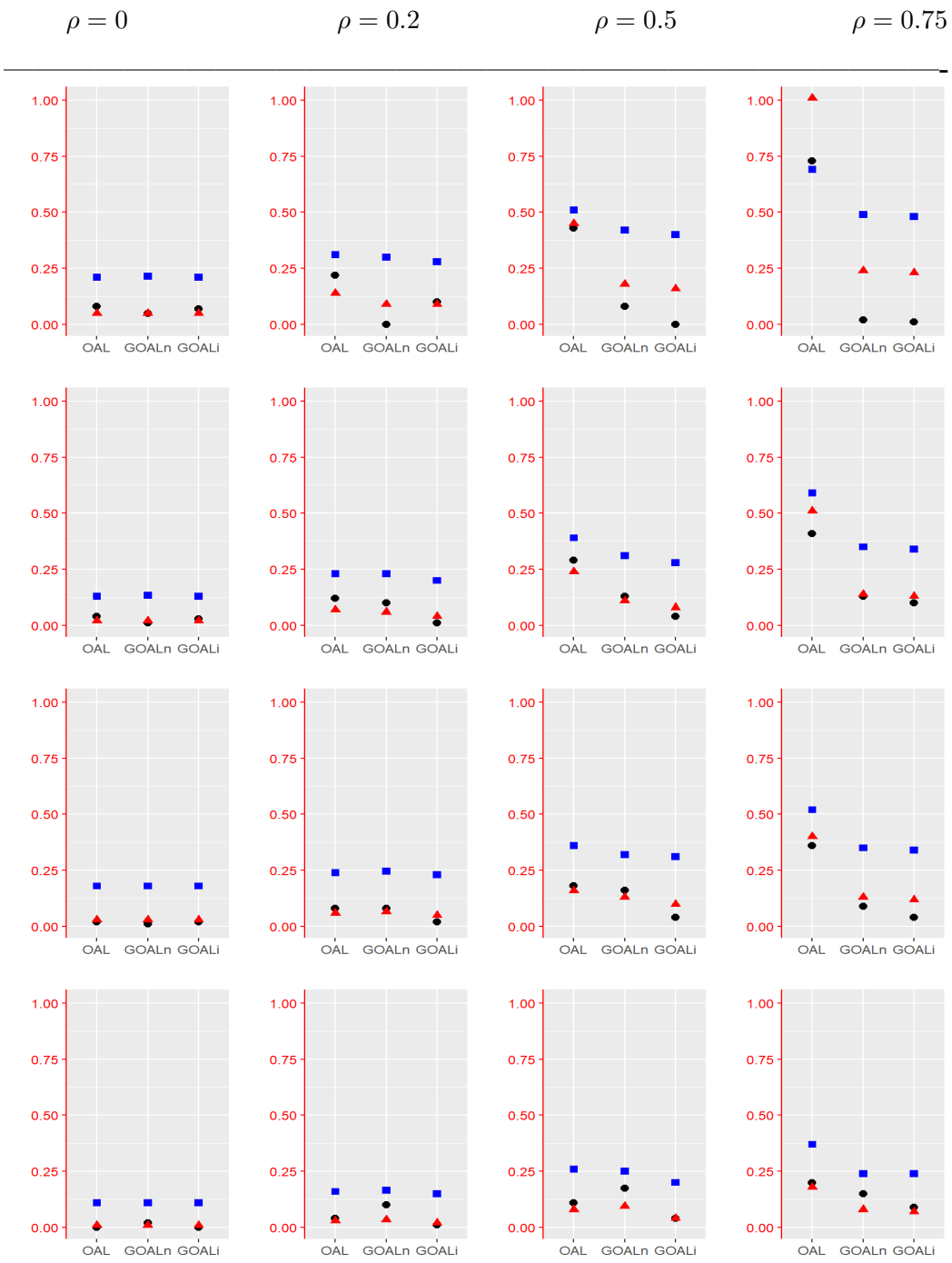
In the low-dimensional settings ($n = 200, 500, 1000$ with $p = 20$), we found that GOALi performed generally much like GOALn for both bias and variance in all scenarios (refer to Web Table 2.2 and Web Figures 2.7-2.10). OAL performed similarly to GOAL estimators when $\rho = 0$. However, when $\rho > 0$, GOAL yielded ATE estimators that were both less biased and less variable than OAL, with the difference between GOAL and OAL becoming more marked as ρ increased. GOAL estimators exhibited small biases for all

ρ values, while OAL's bias increased with correlation.

In Web Figures 2.11-2.14, we present the wAMD (weighted absolute mean difference) between exposure groups for OAL and GOAL estimators over 1000 simulations for combinations ($n = 200$, $p = 100$) and ($n = 200$, $p = 20$) with $\rho = 0, 0.75$. In the high-dimensional setting with $\rho = 0$, OAL and GOALi performed similarly with respect to wAMD values, while GOALn yielded almost systematically larger wAMD values in comparison with GOALi (Web Figure 2.11). In the low-dimensional setting with $\rho = 0$, GOAL produced overall smaller or similar wAMD values than OAL across scenarios (Web Figure 2.13). When $\rho = 0.75$, both settings yielded wAMD values for OAL that were remarkably larger than for GOAL (Web Figure 2.12 and Web Figure 2.14). This greater imbalance between exposure groups for OAL is in line with the larger bias observed for OAL as compared to GOAL in the high- and low- dimensional settings with $\rho = 0.75$.

Presentation of variable selection results (see Web Figures 2.15-2.16) is done in the Web Appendix.

Figure 2.1 Absolute bias (circle), standard error (square) and mean squared error (triangle) of IPTW estimator for the average treatment effect (ATE) for OAL, naive GOAL (GOALn) and penalized iteratively re-weighted least squares GOAL (GOALi) under Scenarios 1 and 2 (based on 1000 IPTW estimates). The ratios $p/n = 100/200, 200/500$ are presented in rows 1 and 2, respectively, for Scenario 1 and in rows 3 and 4, respectively, for Scenario 2. This figure appears in color in the electronic version of this article.



2.6 Discussion

We presented a note on the outcome-adaptive lasso (OAL), a penalized variable selection approach for causal inference proposed by Shortreed and Ertefaie (2017). While OAL was observed to have good performance in the low-dimensional settings with uncorrelated covariates examined by Shortreed and Ertefaie (2017), this approach was found to yield a biased IPTW estimator of the ATE in the high- and low-dimensional settings with correlated covariates. Our proposed approach GOAL was designed to improve on OAL by combining ideas from the adaptive lasso to achieve the oracle property and elastic net to address the collinearity problem. Our results showed that GOAL performed similarly or better than OAL in terms of balance between exposure groups and estimation accuracy in the same simulation scenarios studied by Shortreed and Ertefaie (2017). In particular, both versions of GOAL yielded IPTW estimators that were markedly less biased and variable than OAL in high and low dimensions with strongly correlated covariates. We found that Naive GOAL (GOALn) performed similarly to GOAL with PIRLS (GOALi) in the high-dimensional settings with uncorrelated covariates, but the latter was found less biased and slightly less variable than the former, in general, with correlated covariates. In the low-dimensional settings, GOALn and GOALi performed equivalently for every level of correlation between covariates investigated. Noting that the data augmentation step used for Naive GOALn is not readily justified for logistic models, GOALn's key advantage is that it can be estimated directly with the R code provided by Shortreed and Ertefaie (2017) after the data are modified. Although GOALi is also simple to implement, it is not as convenient as naive GOAL regarding the simplicity of the algorithm.

In this note, we performed all simulations with modest and large number of covariates (p) when $p < n$. A potential extension of this work would be to generalize GOAL for the case $p \geq n$. Indeed, owing to the properties of adaptive elastic net, GOAL appears well equipped for tackling this case. However, GOAL (as well as OAL) is based on the ordinary least squares (ols) weights $\hat{w}$ which require a full rank model for their estimation.

GOAL's extension to ultra high dimension will thus necessitate some modification to how the adaptive weights are defined. This will be investigated in a future study.

2.7 Web Appendix

2.7.1 Web Appendix A. Adaptive elastic net

Elastic net performs simultaneous regularization and variable selection, with a penalty that is a convex combination of the lasso and ridge penalties (Zou and Hastie, 2005). Compared to lasso which selects at most n variables, elastic net is viewed to be better equiped to tackle the $p \gg n$ problem and to deal with grouped selection of correlated variables. Indeed, with correlated features, lasso is known to exhibit variable selection problems as it tends to select only one variable from the group (Zou and Hastie, 2005). The adaptive elastic net method was specifically introduced to analyse high-dimensional data (Zou and Zhang, 2009). It is a mixture of adaptive lasso and elastic net, inheriting good properties from each: the oracle property from the former and the ability to overcome the collinearity problem from the latter (Zou and Zhang, 2009; Ghosh, 2011). Adaptive elastic net has also shown advantages in practice, for example in high-dimensional cancer classification for simultaneous estimation and gene selection (Algamal and Lee, 2015b).

Consider a standard linear regression problem along with nonnegative tuning parameters (λ_1, λ_2). Ghosh (2011) defined the adaptive elastic net estimator as

$$\hat{\beta} = \arg\min_{\beta} \left(\|Y - \mathbf{X}\beta\|_2^2 + \lambda_1 \sum_{j=1}^{p} \tilde{w}_j |\beta_j| + \lambda_2 \sum_{j=1}^{p} \beta_j^2 \right), \tag{2.8}$$

where $\tilde{w}_j = \left|\tilde{\beta}_j\right|^{-\gamma}$ such that $\gamma > 0$ and $\tilde{\beta}$ is a root n-consistent estimator of β.
Similar to elastic net (Zou and Hastie, 2005), Ghosh (2011) showed the equivalence between adaptive elastic net (that is, Equation (2.8)) and an ordinary adaptive lasso

problem in some augmented space:

$$\hat{\beta} = \arg\min_{\beta} \left(\|Y^* - \mathbf{X}^*\beta\|_2^2 + \lambda_1 \sum_{j=1}^{p} \tilde{w}_j |\beta_j| \right), \tag{2.9}$$

where $\mathbf{X}^* = \begin{pmatrix} \mathbf{X} \\ \sqrt{\lambda_2}\mathbf{I}_p \end{pmatrix}$, $Y^* = \begin{pmatrix} Y \\ 0_p \end{pmatrix}$.
Consequently, $\mathbf{X}^*$ has rank p and adaptive elastic net can potentially select all p predictors when either $p < n$ or $p \geqslant n$.

In what follows, we recall the data augmentation proof of the adaptive elastic net (Ghosh, 2011). Let $\mathbf{I}_p$ be a $p \times p$ identity matrix and $0_p = (0, 0, \ldots, 0)^T \in \mathbb{R}^p$.

For any fixed λ_2 we have:

$$\begin{aligned}
&\left\| \; Y - \mathbf{X}\beta \; \right\|_2^2 + \lambda_2 \left\| \; \beta \; \right\|_2^2 = \left\| \; Y - \mathbf{X}\beta \; \right\|_2^2 + \left\| \; 0_p - \sqrt{\lambda_2}\mathbf{I}_p\beta \; \right\|_2^2 \\
&= \left(\; Y - \mathbf{X}\beta \; \right)^T \left(\; Y - \mathbf{X}\beta \; \right) + \left(\; 0_p - \sqrt{\lambda_2}\mathbf{I}_p\beta \; \right)^T \left(\; 0_p - \sqrt{\lambda_2}\mathbf{I}_p\beta \; \right) \\
&= \begin{pmatrix} Y - \mathbf{X}\beta \\ 0_p - \sqrt{\lambda_2}\mathbf{I}_p\beta \end{pmatrix}^T \begin{pmatrix} Y - \mathbf{X}\beta \\ 0_p - \sqrt{\lambda_2}\mathbf{I}_p\beta \end{pmatrix} = \left\| \begin{matrix} Y - \mathbf{X}\beta \\ 0_p - \sqrt{\lambda_2}\mathbf{I}_p\beta \end{matrix} \right\|_2^2 \\
&= \left\| \; Y^* - \mathbf{X}^*\beta \; \right\|_2^2, \quad Y^* = \begin{pmatrix} Y \\ 0_p \end{pmatrix}, \mathbf{X}^* = \begin{pmatrix} \mathbf{X} \\ \sqrt{\lambda_2}\mathbf{I}_p \end{pmatrix}.
\end{aligned}$$

Thus Equation (2.8) becomes Equation (2.9)

$$\hat{\beta} = \arg\min_{\beta} \left(\|Y^* - \mathbf{X}^*\beta\|_2^2 + \lambda_1 \sum_{j=1}^{p} \tilde{w}_j |\beta_j| \right).$$

2.7.2 Web Appendix B. Data augmentation for the proposed penalized iteratively re-weighted least squares GOAL (GOALi)

The GOAL estimator is defined as

$$\hat{\alpha}(GOAL) = \arg\min_{\alpha} \left[\ell_n(\alpha; A, \mathbf{X}) + \lambda_1 \sum_{j=1}^{p} \hat{w}_j |\alpha_j| + \lambda_2 \sum_{j=1}^{p} \alpha_j^2 \right],$$

where $\hat{w}_j = \left|\hat{\beta}_j^{ols}\right|^{-\gamma}$ such that $\gamma > 1$ and $(\hat{\beta}_A^{ols}, \hat{\beta}^{ols}) = \arg\min_{(\beta_A,\beta)} \|Y - \beta_A A - \mathbf{X}\beta\|_2^2$. The Newton-Raphson update solution of GOAL is obtained as

$$\hat{\alpha}_{PIRLS}(GOAL) = \arg\min_{\alpha} \left[\ell_Q(\alpha; A, \mathbf{X}, Z, \mathbf{T}) + \lambda_1 \sum_{j=1}^{p} \hat{w}_j |\alpha_j| + \lambda_2 \sum_{j=1}^{p} \alpha_j^2 \right], \quad (2.10)$$

where $\ell_Q(\alpha; A, \mathbf{X}, Z, \mathbf{T}) = \frac{1}{2}\sum_{i=1}^{n} t_i \left(z_i - x_i^T\alpha\right)^2$, $\quad \tilde{p}(x_i) = \frac{1}{1+\exp(-x_i^T\tilde{\alpha})}$,

$t_i = \tilde{p}(x_i)[1-\tilde{p}(x_i)]$, $\quad z_i = x_i^T\tilde{\alpha} + \frac{a_i - \tilde{p}(x_i)}{\tilde{p}(x_i)(1-\tilde{p}(x_i))}$, $\quad Z = (z_1, \ldots, z_n)^T$, $\quad \mathbf{T} = (t_1, \ldots, t_n)$. The data augmentation step of the proposed GOAL with PIRLS method (GOALi) is obtained as follows. Let $0_{n\times p}$ be a $n \times p$ null matrix. For any fixed λ_2, we have:

$$\left\| \mathbf{T}^{1/2}(Z - \mathbf{X}\alpha) \right\|_2^2 + \lambda_2 \left\| \alpha \right\|_2^2 = \left\| \mathbf{T}^{1/2}(Z - \mathbf{X}\alpha) \right\|_2^2 + \left\| \mathbf{I}_p^{1/2}\left(0_p - \sqrt{\lambda_2}\mathbf{I}_p\alpha\right) \right\|_2^2$$

$$= \left\| \mathbf{T}^{1/2}(Z - \mathbf{X}\alpha) + 0_{n\times p}\left(0_p - \sqrt{\lambda_2}\mathbf{I}_p\alpha\right) \right\|_2^2 + \left\| 0_{n\times p}^T(Z - \mathbf{X}\alpha) + \mathbf{I}_p^{1/2}\left(0_p - \sqrt{\lambda_2}\mathbf{I}_p\alpha\right) \right\|_2^2$$

$$= \left\| \left(\mathbf{T}^{1/2}, 0_{n\times p}\right) \begin{pmatrix} Z - \mathbf{X}\alpha \\ 0_p - \sqrt{\lambda_2}\mathbf{I}_p\alpha \end{pmatrix} \right\|_2^2 + \left\| \left(0_{n\times p}^T, \mathbf{I}_p^{1/2}\right) \begin{pmatrix} Z - \mathbf{X}\alpha \\ 0_p - \sqrt{\lambda_2}\mathbf{I}_p\alpha \end{pmatrix} \right\|_2^2$$

$$= \left\| \begin{pmatrix} \mathbf{T}^{1/2} & 0_{n\times p} \\ 0_{n\times p}^T & \mathbf{I}_p^{1/2} \end{pmatrix} \begin{pmatrix} Z - \mathbf{X}\alpha \\ 0_p - \sqrt{\lambda_2}\mathbf{I}_p\alpha \end{pmatrix} \right\|_2^2 = \left\| \mathbf{T}^{*1/2}(Z^* - \mathbf{X}^*\alpha) \right\|_2^2,$$

$$Z^* = \begin{pmatrix} Z \\ 0_p \end{pmatrix}, \quad A^* = \begin{pmatrix} A \\ 0_p \end{pmatrix}, \quad \mathbf{X}^* = \begin{pmatrix} \mathbf{X} \\ \sqrt{\lambda_2}\mathbf{I}_p \end{pmatrix} \text{ and } \quad \mathbf{T}^{*1/2} = \begin{pmatrix} \mathbf{T}^{1/2} & 0_{n\times p} \\ 0_{n\times p}^T & \mathbf{I}_p^{1/2} \end{pmatrix}.$$

Thus Equation (2.10) becomes

$$\hat{\alpha}_Q(GOAL) = \arg\min_{\alpha} \left[\ell_{Q^*}(\alpha; A^*, \mathbf{X}^*, Z^*, \mathbf{T}^*) + \lambda_1 \sum_{j=1}^{p} \hat{w}_j |\alpha_j| \right],$$

where

$$\ell_{Q^*}(\alpha; A^*, \mathbf{X}^*, Z^*, \mathbf{T}^*) = \left\| \mathbf{T}^{*1/2}(Z^* - \mathbf{X}^*\alpha) \right\|_2^2.$$

2.7.3 Web Appendix C. Simulation Results

Web Table 2.1 presents the bias, standard error (SE) and mean squared error (MSE) for OAL and GOAL estimators under all scenarios in the high-dimensional settings ($p/n = 100/200, 200/500$).

Web Table 2.2 presents the bias, SE and MSE for OAL and GOAL estimators under all scenarios in the low-dimensional settings ($n = 200, 500, 1000$ and $p = 20$).

Tableau 2.1 Bias (SE; **MSE**) of the IPTW estimator for the average treatment effect (ATE) for OAL, GOALn and GOALi with ratios $p/n = 100/200, 200/500$ under Scenarios 1, 2, 3 and 4 by sections 1, 2, 3 and 4, respectively (results based on 1000 estimates of the ATE).

	$\frac{p}{n}$		$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.75$
1		OAL	0.08 (0.21; **0.05**)	0.22 (0.31; **0.14**)	0.43 (0.51; **0.45**)	0.73 (0.69; **1.01**)
	$\frac{100}{200}$	GOALn	0.05 (0.22; **0.05**)	0.00 (0.30; **0.09**)	-0.08 (0.42; **0.18**)	-0.02 (0.49; **0.24**)
		GOALi	0.07 (0.21; **0.05**)	0.10 (0.28; **0.09**)	0.00 (0.40; **0.16**)	0.01 (0.48; **0.23**)
		OAL	0.04 (0.13; **0.02**)	0.12 (0.23; **0.07**)	0.29 (0.39; **0.24**)	0.41 (0.59; **0.51**)
	$\frac{200}{500}$	GOALn	0.01 (0.14; **0.02**)	-0.10 (0.23; **0.06**)	-0.13 (0.31; **0.11**)	-0.13 (0.35; **0.14**)
		GOALi	0.03 (0.13; **0.02**)	0.01 (0.20; **0.04**)	-0.04 (0.28; **0.08**)	-0.10 (0.34; **0.13**)
2		OAL	0.02 (0.18; **0.03**)	0.08 (0.24; **0.06**)	0.18 (0.36; **0.16**)	0.36 (0.52; **0.40**)
	$\frac{100}{200}$	GOALn	-0.01 (0.18; **0.03**)	-0.08 (0.25; **0.07**)	-0.16 (0.32; **0.13**)	-0.09 (0.35; **0.13**)
		GOALi	0.02 (0.18; **0.03**)	0.02 (0.23; **0.05**)	-0.04 (0.31; **0.10**)	-0.04 (0.34; **0.12**)
		OAL	0.00 (0.11; **0.01**)	0.04 (0.16; **0.03**)	0.11 (0.26; **0.08**)	0.20 (0.37; **0.18**)
	$\frac{200}{500}$	GOALn	-0.02 (0.11; **0.01**)	-0.10 (0.17; **0.04**)	-0.18 (0.25; **0.10**)	-0.15 (0.24; **0.08**)
		GOALi	0.00 (0.11; **0.01**)	-0.01 (0.15; **0.02**)	-0.04 (0.20; **0.04**)	-0.09 (0.24; **0.07**)
3		OAL	0.09 (0.20; **0.05**)	0.16 (0.26; **0.09**)	0.30 (0.40; **0.25**)	0.48 (0.52; **0.50**)
	$\frac{100}{200}$	GOALn	0.12 (0.20; **0.06**)	0.02 (0.29; **0.08**)	-0.05 (0.39; **0.15**)	-0.05 (0.46; **0.21**)
		GOALi	0.09 (0.21; **0.05**)	0.10 (0.27; **0.08**)	0.03 (0.38; **0.14**)	-0.01 (0.45; **0.21**)
		OAL	0.04 (0.13; **0.02**)	0.07 (0.19; **0.04**)	0.20 (0.31; **0.14**)	0.29 (0.46; **0.29**)
	$\frac{200}{500}$	GOALn	0.06 (0.15; **0.03**)	-0.07 (0.21; **0.05**)	-0.13 (0.33; **0.13**)	-0.15 (0.34; **0.14**)
		GOALi	0.04 (0.13; **0.02**)	0.01 (0.19; **0.03**)	-0.01 (0.27; **0.07**)	-0.09 (0.35; **0.13**)
4		OAL	0.05 (0.19; **0.04**)	0.19 (0.30; **0.12**)	0.52 (0.52; **0.55**)	1.03 (0.70; **1.54**)
	$\frac{100}{200}$	GOALn	0.01 (0.20; **0.04**)	-0.06 (0.32; **0.11**)	-0.12 (0.49; **0.26**)	-0.03 (0.57; **0.33**)
		GOALi	0.05 (0.19; **0.04**)	0.06 (0.29; **0.09**)	-0.01 (0.46; **0.21**)	-0.04 (0.57; **0.33**)
		OAL	0.02 (0.11; **0.01**)	0.14 (0.22; **0.07**)	0.42 (0.41; **0.34**)	0.68 (0.69; **0.93**)
	$\frac{200}{500}$	GOALn	-0.02 (0.12; **0.01**)	-0.11 (0.23; **0.07**)	-0.17 (0.36; **0.16**)	-0.20 (0.46; **0.25**)
		GOALi	0.01 (0.11; **0.01**)	0.00 (0.20; **0.04**)	-0.04 (0.34; **0.11**)	-0.17 (0.44; **0.23**)

Tableau 2.2: Bias (SE; **MSE**) of the IPTW estimator for the average treatment effect (ATE) for OAL, GOALn and GOALi with fixed $p = 20$ and increasing $n = 200, 500, 1000$ under Scenarios 1, 2, 3 and 4 by sections 1, 2, 3 and 4, respectively (results based on 1000 estimates of the ATE).

	n		$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.75$
1		OAL	0.04 (0.19; **0.04**)	0.15 (0.28; **0.10**)	0.41 (0.44; **0.36**)	0.65 (0.59; **0.78**)
	200	GOALn	0.01 (0.18; **0.03**)	0.03 (0.25; **0.06**)	0.03 (0.33; **0.11**)	0.02 (0.44; **0.20**)
		GOALi	0.01 (0.18; **0.03**)	0.02 (0.24; **0.06**)	0.03 (0.32; **0.10**)	0.02 (0.42; **0.18**)
		OAL	0.03 (0.12; **0.02**)	0.12 (0.17; **0.04**)	0.31 (0.32; **0.20**)	0.54 (0.47; **0.51**)
	500	GOALn	0.01 (0.12; **0.01**)	0.02 (0.15; **0.02**)	0.01 (0.22; **0.05**)	0.01 (0.29; **0.08**)
		GOALi	0.01 (0.12; **0.01**)	0.02 (0.15; **0.02**)	0.00 (0.22; **0.05**)	0.01 (0.29; **0.08**)
		OAL	0.01 (0.08; **0.01**)	0.09 (0.13; **0.02**)	0.26 (0.26; **0.14**)	0.46 (0.40; **0.37**)
	1000	GOALn	0.00 (0.08; **0.01**)	0.01 (0.10; **0.01**)	-0.01 (0.17; **0.03**)	-0.01 (0.23; **0.05**)
		GOALi	0.00 (0.08; **0.01**)	0.01 (0.10; **0.01**)	-0.01 (0.17; **0.03**)	-0.01 (0.23; **0.05**)
2		OAL	0.00 (0.15; **0.02**)	0.04 (0.20; **0.04**)	0.18 (0.30; **0.12**)	0.35 (0.44; **0.32**)
	200	GOALn	-0.01 (0.15; **0.02**)	-0.01 (0.19; **0.04**)	-0.01 (0.24; **0.06**)	0.00 (0.31; **0.10**)
		GOALi	-0.01 (0.15; **0.02**)	0.00 (0.19; **0.04**)	-0.01 (0.24; **0.06**)	0.01 (0.30; **0.09**)
		OAL	0.01 (0.11; **0.01**)	0.03 (0.12; **0.01**)	0.11 (0.21; **0.06**)	0.25 (0.31; **0.16**)
	500	GOALn	0.01 (0.11; **0.01**)	0.00 (0.11; **0.01**)	-0.02 (0.16; **0.02**)	-0.01 (0.20; **0.04**)
		GOALi	0.01 (0.11; **0.01**)	0.00 (0.11; **0.01**)	-0.02 (0.16; **0.02**)	-0.01 (0.20; **0.04**)
		OAL	0.00 (0.07; **0.00**)	0.02 (0.08; **0.01**)	0.10 (0.15; **0.03**)	0.19 (0.25; **0.10**)
	1000	GOALn	0.00 (0.07; **0.00**)	0.00 (0.08; **0.01**)	-0.01 (0.11; **0.01**)	-0.02 (0.14; **0.02**)
		GOALi	0.00 (0.07; **0.00**)	0.00 (0.08; **0.01**)	-0.01 (0.11; **0.01**)	-0.02 (0.14; **0.02**)

Continued on next page

Tableau 2.2 – *Continued from previous page*

	n		$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.75$
3		OAL	0.04 (0.19; **0.04**)	0.10 (0.25; **0.07**)	0.25 (0.32; **0.17**)	0.40 (0.44; **0.35**)
	200	GOALn	0.03 (0.19; **0.04**)	0.03 (0.25; **0.06**)	0.04 (0.31; **0.10**)	0.01 (0.41; **0.17**)
		GOALi	0.03 (0.19; **0.04**)	0.03 (0.24; **0.06**)	0.04 (0.30; **0.09**)	0.03 (0.40; **0.16**)
		OAL	0.01 (0.12; **0.01**)	0.07 (0.16; **0.03**)	0.19 (0.27; **0.11**)	0.33 (0.34; **0.22**)
	500	GOALn	0.01 (0.12; **0.01**)	0.01 (0.15; **0.02**)	-0.01 (0.23; **0.05**)	0.01 (0.28; **0.08**)
		GOALi	0.01 (0.12; **0.01**)	0.01 (0.15; **0.02**)	-0.01 (0.23; **0.05**)	0.01 (0.28; **0.08**)
		OAL	0.00 (0.08; **0.01**)	0.04 (0.12; **0.02**)	0.16 (0.21; **0.07**)	0.29 (0.28; **0.16**)
	1000	GOALn	0.00 (0.08; **0.01**)	-0.01 (0.11; **0.01**)	-0.02 (0.18; **0.03**)	0.00 (0.22; **0.05**)
		GOALi	0.00 (0.08; **0.01**)	-0.01 (0.11; **0.01**)	-0.02 (0.17; **0.03**)	0.00 (0.22; **0.05**)
4		OAL	0.01 (0.17; **0.03**)	0.14 (0.26; **0.09**)	0.46 (0.44; **0.41**)	0.92 (0.62; **1.23**)
	200	GOALn	0.00 (0.17; **0.03**)	0.03 (0.24; **0.06**)	0.03 (0.36; **0.13**)	0.03 (0.52; **0.28**)
		GOALi	0.00 (0.17; **0.03**)	0.03 (0.24; **0.06**)	0.03 (0.36; **0.13**)	0.02 (0.51; **0.26**)
		OAL	0.02 (0.11; **0.01**)	0.09 (0.15; **0.03**)	0.38 (0.31; **0.24**)	0.77 (0.52; **0.86**)
	500	GOALn	0.01 (0.11; **0.01**)	0.01 (0.14; **0.02**)	0.01 (0.23; **0.06**)	0.00 (0.35; **0.12**)
		GOALi	0.01 (0.11; **0.01**)	0.01 (0.14; **0.02**)	0.01 (0.23; **0.05**)	0.00 (0.35; **0.12**)
		OAL	0.01 (0.07; **0.01**)	0.08 (0.12; **0.02**)	0.34 (0.25; **0.18**)	0.72 (0.42; **0.70**)
	1000	GOALn	0.00 (0.07; **0.01**)	0.01 (0.10; **0.01**)	0.00 (0.17; **0.03**)	-0.01 (0.27; **0.08**)
		GOALi	0.00 (0.07; **0.01**)	0.01 (0.10; **0.01**)	0.00 (0.17; **0.03**)	-0.01 (0.27; **0.08**)

Web Figure 2.2 displays the bias, SE and MSE for OAL and GOAL estimators under Scenarios 3 and 4 in the high-dimensional settings ($p/n = 100/200, 200/500$).

Web Figures 2.3, 2.4, 2.5 and 2.6 present the box plots of ATE estimates for OAL and GOAL estimators under all scenarios in the high-dimensional settings ($p/n = 100/200, 200/500$) for $\rho = 0$, 0.2, 0.5 and 0.75, respectively.

Figure 2.2 Absolute bias (circle), standard error (square) and mean squared error (triangle) of IPTW estimator for the average treatment effect (ATE) for OAL, naive GOAL (GOALn) and penalized iteratively re-weighted least squares GOAL (GOALi) under Scenarios 3 and 4 (based on 1000 IPTW estimates). The ratios $p/n = 100/200, 200/500$ are presented in rows 1 and 2, respectively, for Scenario 3 and in rows 3 and 4, respectively, for Scenario 4.

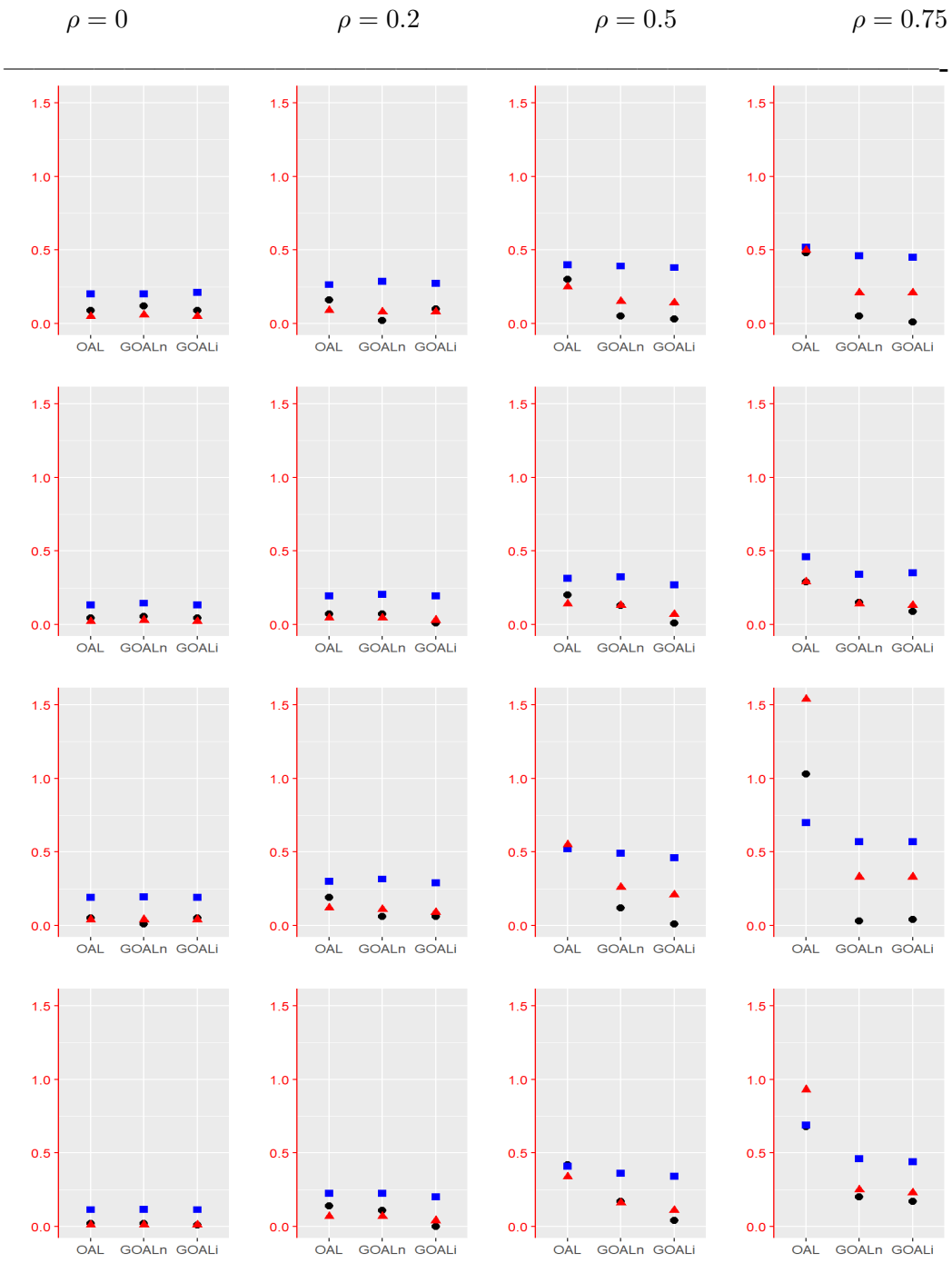


Figure 2.3 Box plots of 1000 IPTW estimates for the average treatment effect (ATE) for OAL, naive GOAL (GOALn) and penalized iteratively re-weighted least squares GOAL (GOALi) under Scenarios 1, 2, 3 and 4 (by row) in the high-dimensional settings with $\rho = 0$. The true value of ATE is indicated with dotted line (ATE=0).

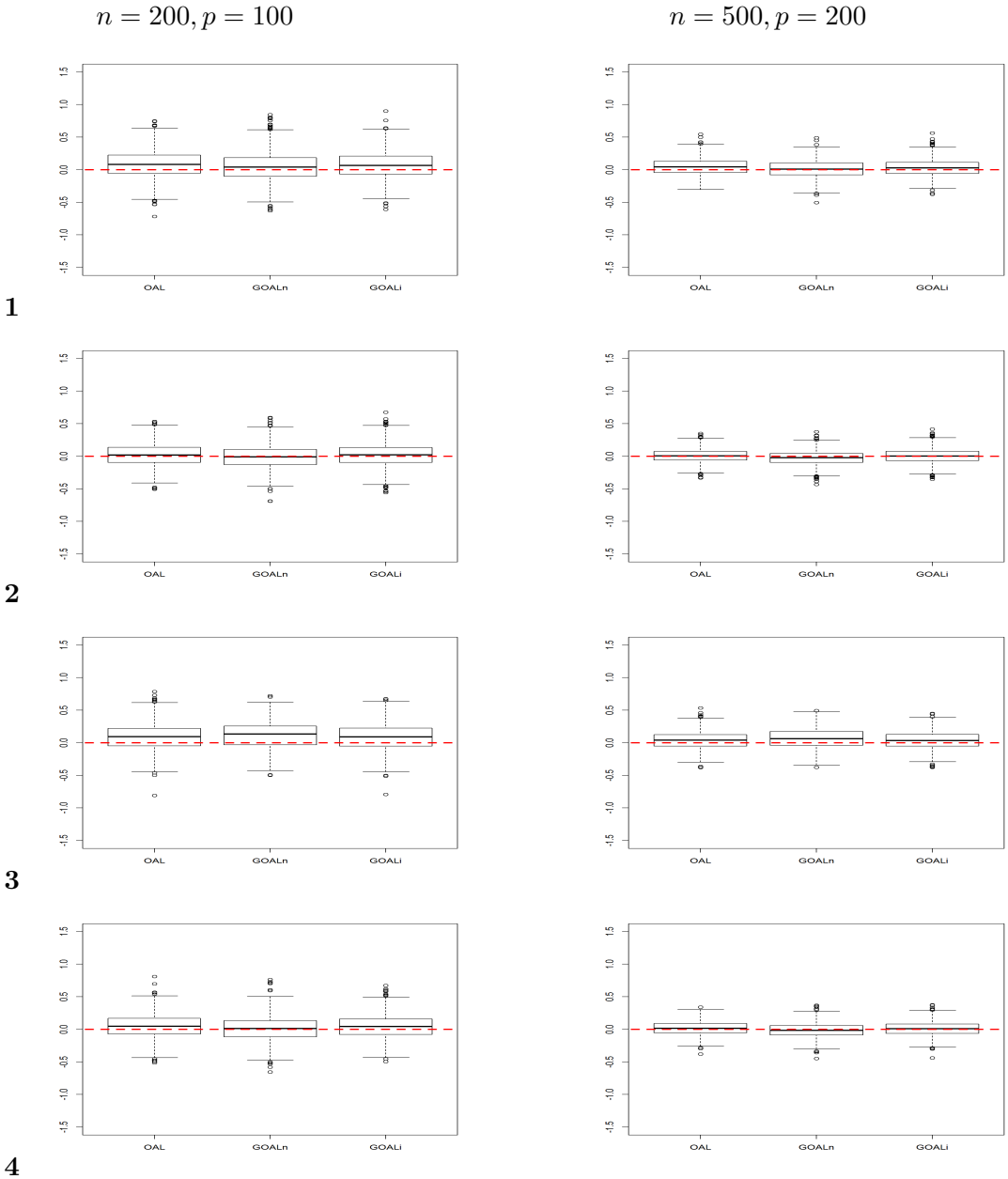


Figure 2.4 Box plots of 1000 IPTW estimates for the average treatment effect (ATE) for OAL, naive GOAL (GOALn) and penalized iteratively re-weighted least squares GOAL (GOALi) under Scenarios 1, 2, 3 and 4 (by row) in the high-dimensional settings with $\rho = 0.2$. The true value of ATE is indicated with dotted line (ATE=0).

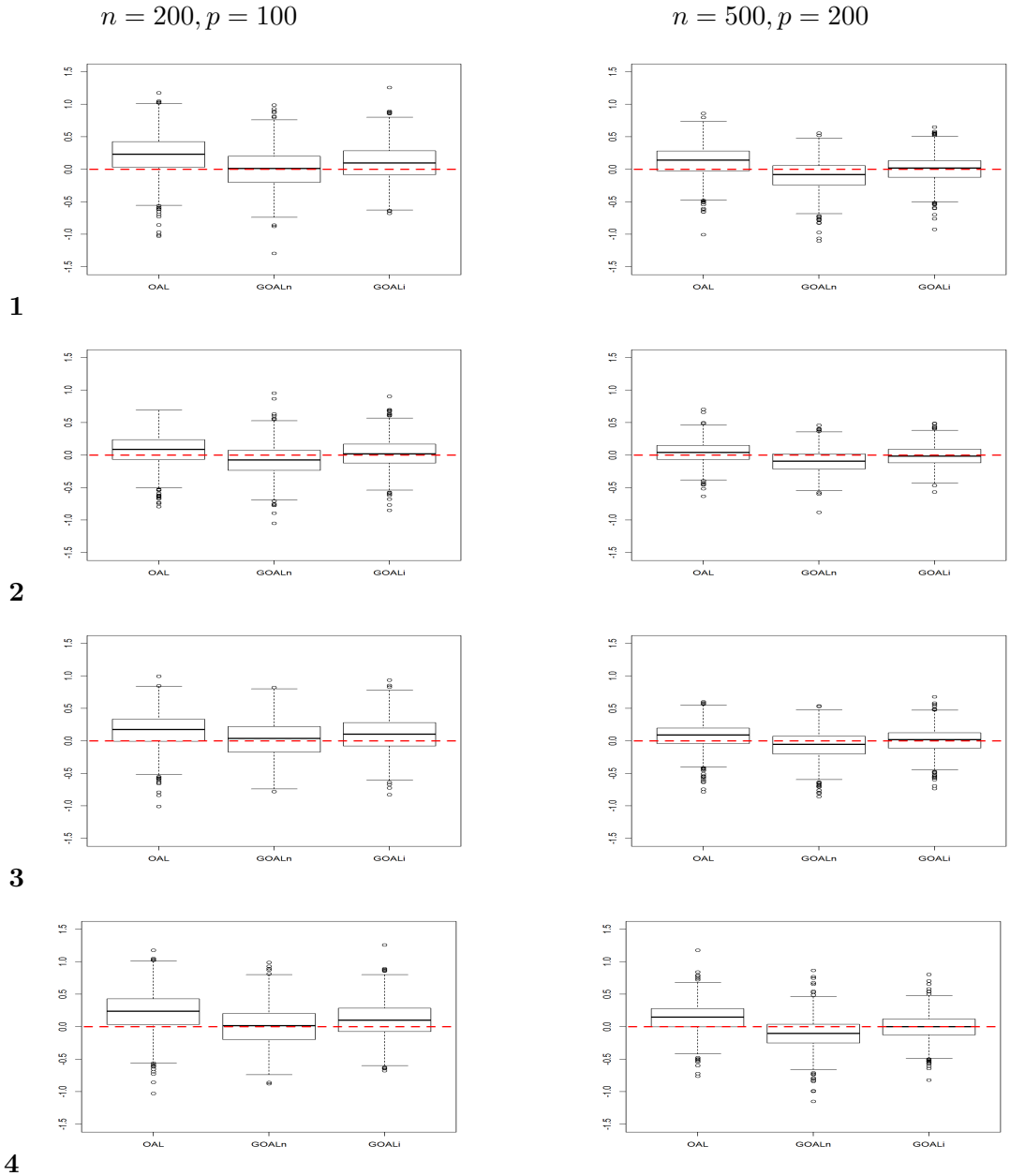


Figure 2.5 Box plots of 1000 IPTW estimates for the average treatment effect (ATE) for OAL, naive GOAL (GOALn) and penalized iteratively re-weighted least squares GOAL (GOALi) under Scenarios 1, 2, 3 and 4 (by row) in the high-dimensional settings with $\rho = 0.5$. The true value of ATE is indicated with dotted line (ATE=0).

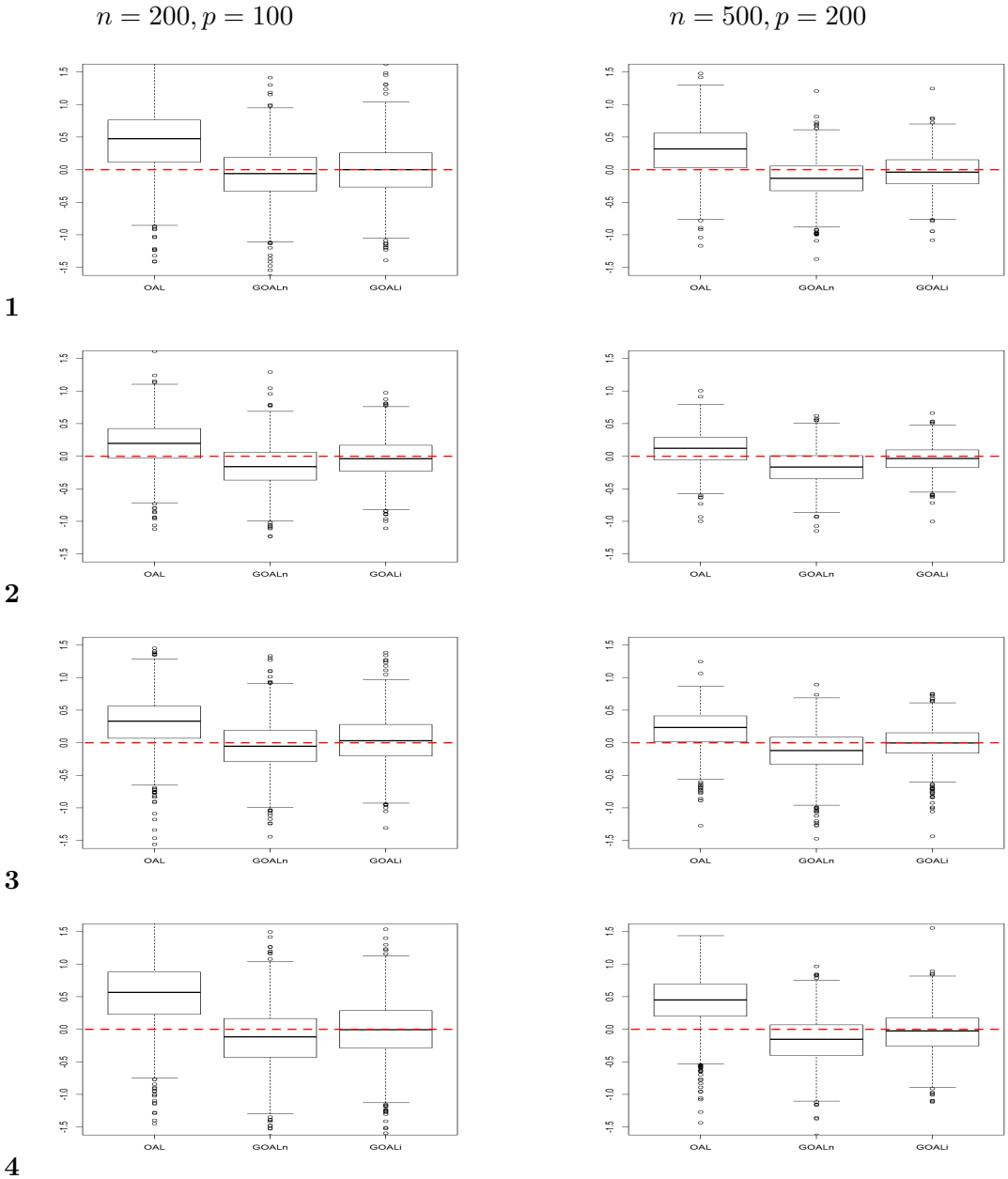
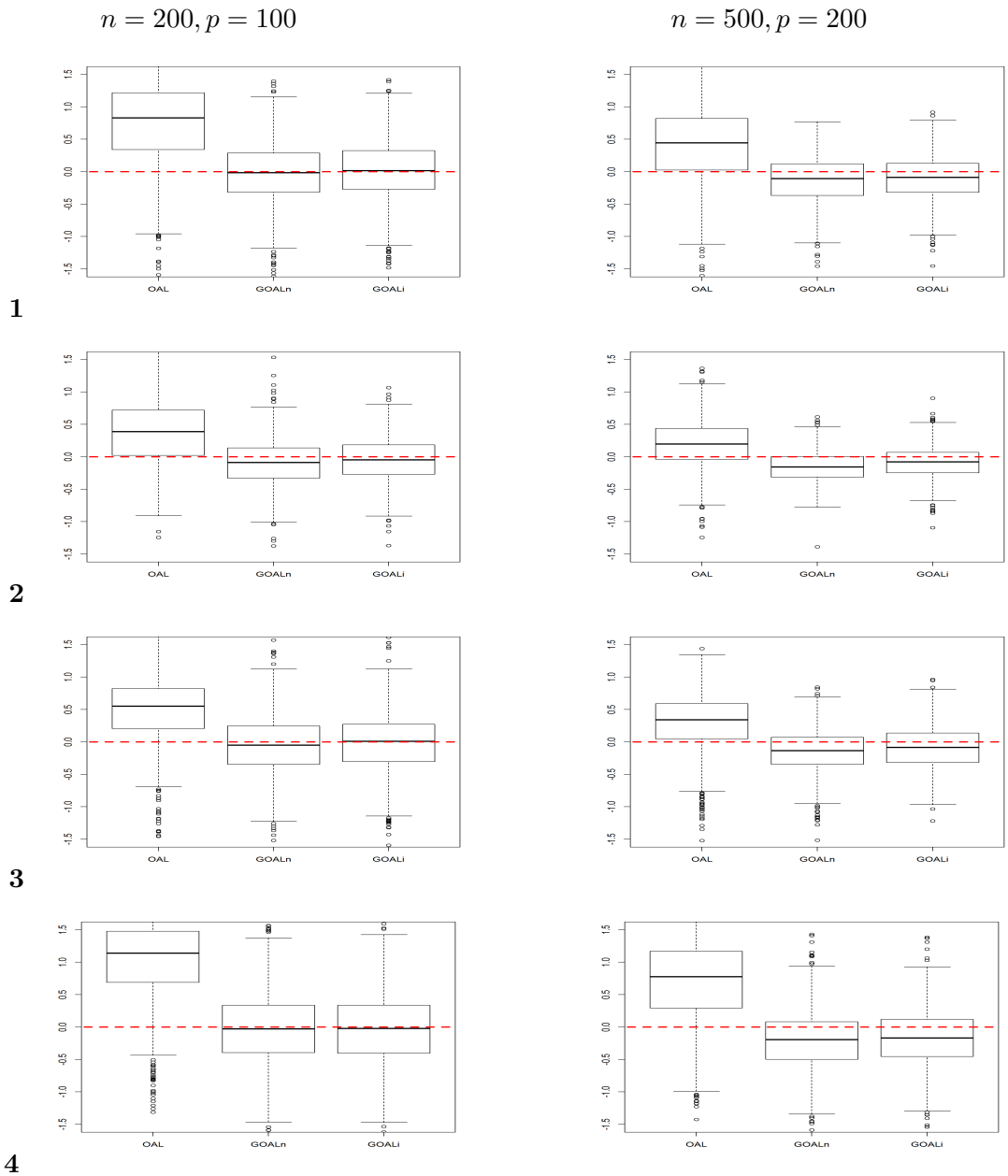


Figure 2.6 Box plots of 1000 IPTW estimates for the average treatment effect (ATE) for OAL, naive GOAL (GOALn) and penalized iteratively re-weighted least squares GOAL (GOALi) under Scenarios 1, 2, 3 and 4 (by row) in the high-dimensional settings with $\rho = 0.75$. The true value of ATE is indicated with dotted line (ATE=0).



Web Figures 2.7, 2.8, 2.9 and 2.10 present the box plots of ATE estimates for OAL and GOAL estimators under all scenarios in the low-dimensional settings ($n = 200, 500, 1000$ with $p = 20$) for $\rho = 0$, 0.2, 0.5 and 0.75, respectively.

Web Figures 2.11-2.14 present the wAMD (weighted absolute mean difference) between exposure groups for OAL and GOAL estimators over 1000 simulations. Results for combination ($n = 200$, $p = 100$) are displayed in Web Figures 2.11 and 2.12 for $\rho = 0$ and 0.75, respectively, while those for the combination ($n = 200$, $p = 20$) appear in Web Figures 2.13 and 2.14 for $\rho = 0$ and 0.75, respectively.

Web Figures 2.15-2.16 show the proportion of times each covariate was selected over 1000 simulations for inclusion in the PS model for combinations ($n = 200$, $p = 100$) and ($n = 200$, $p = 20$) with $\rho = 0, 0.75$. In the high-dimensional setting ($n = 200$, $p = 100$), all estimators (OAL, GOALn and GOALi) included confounders and predictors of the outcome at similar rates. In this setting and for both correlation values ($\rho = 0, 0.75$), GOALn excluded more pure predictors of the exposure and spurious covariates than OAL. In all scenarios, GOALi included more of these variables than OAL when $\rho = 0$, while the same phenomenon was only observed in Scenario 4 when $\rho = 0.75$. In the low-dimensional setting with $\rho = 0$, OAL and GOAL included all covariates at very similar rates. For the low-dimensional setting with $\rho = 0.75$, GOAL included more or slightly more pure predictors of the exposure and spurious covariates than OAL. Moreover, the confounders and the predictors of the outcome were selected by OAL and GOAL at similar rates, except for Scenario 3 for which GOAL noticeably selected more confounders than OAL. In both high- and low-dimensional settings, the number of covariates selected by the estimators was greater, on average, when $\rho = 0.75$ than when $\rho = 0$. Moreover, for a fixed ρ value, the number of covariates selected by OAL and GOAL was far greater in the high-dimensional setting as compared to the low-dimensional setting.

Figure 2.7 Box plots of 1000 IPTW estimates for the average treatment effect (ATE) for OAL, naive GOAL (GOALn) and penalized iteratively re-weighted least squares GOAL (GOALi) under Scenarios 1, 2, 3 and 4 (by row) in the low-dimensional settings with $\rho = 0$. The true value of ATE is indicated with dotted line (ATE=0).

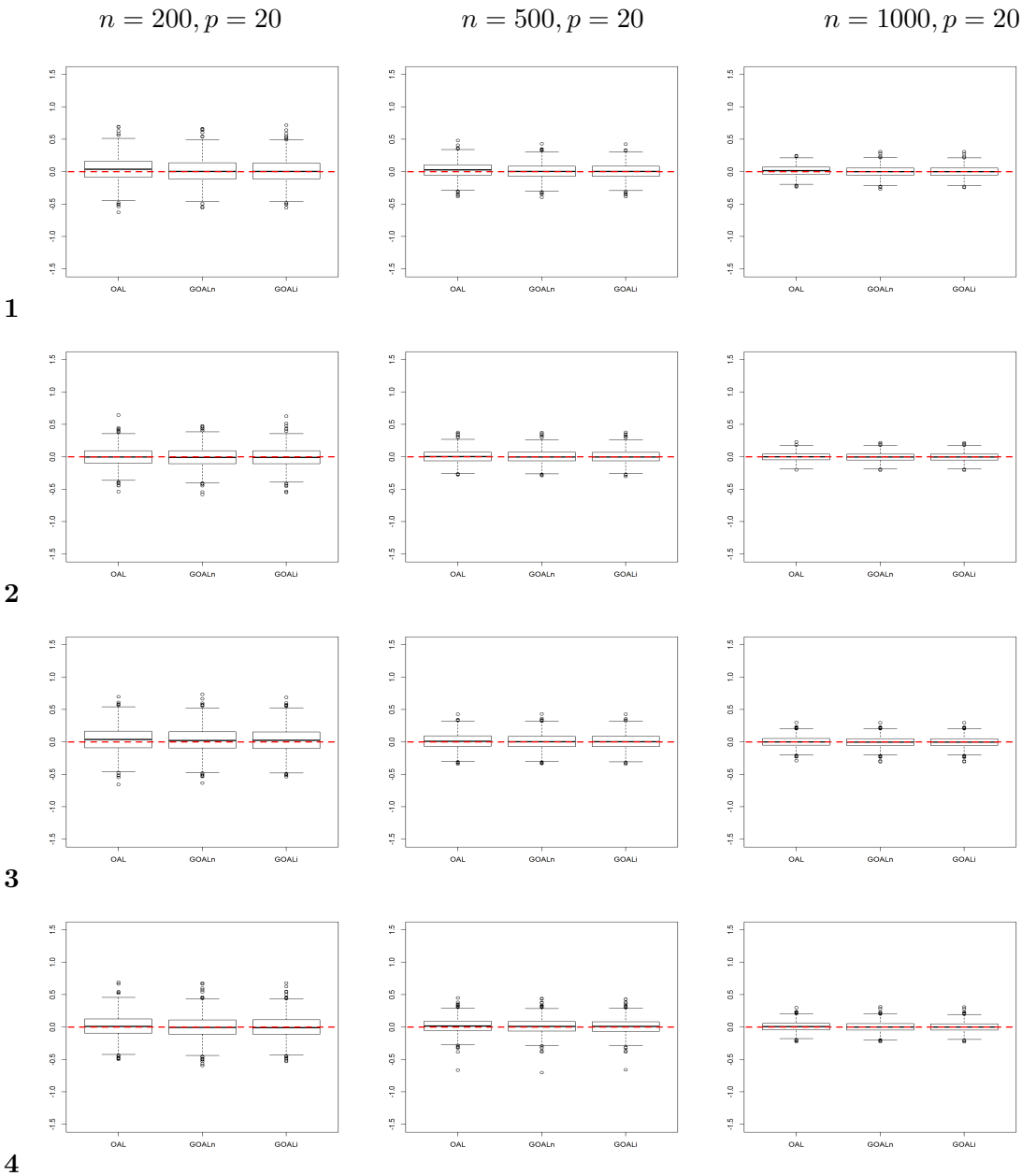


Figure 2.8 Box plots of 1000 IPTW estimates for the average treatment effect (ATE) for OAL, naive GOAL (GOALn) and penalized iteratively re-weighted least squares GOAL (GOALi) under Scenarios 1, 2, 3 and 4 (by row) in the low-dimensional settings with $\rho = 0.2$. The true value of ATE is indicated with dotted line (ATE=0).

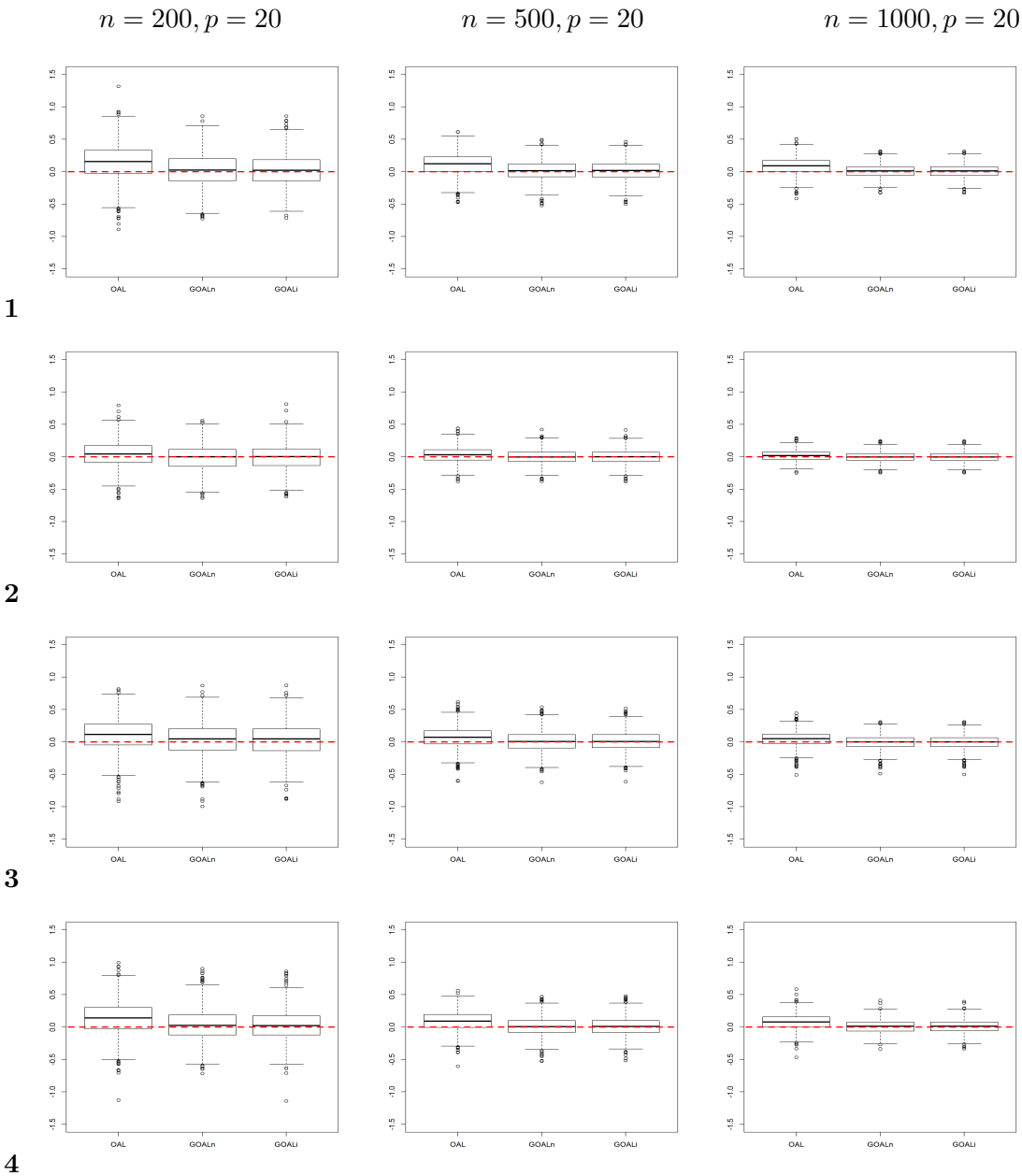


Figure 2.9 Box plots of 1000 IPTW estimates for the average treatment effect (ATE) for OAL, naive GOAL (GOALn) and penalized iteratively re-weighted least squares GOAL (GOALi) under Scenarios 1, 2, 3 and 4 (by row) in the low-dimensional settings with $\rho = 0.5$. The true value of ATE is indicated with dotted line (ATE=0).

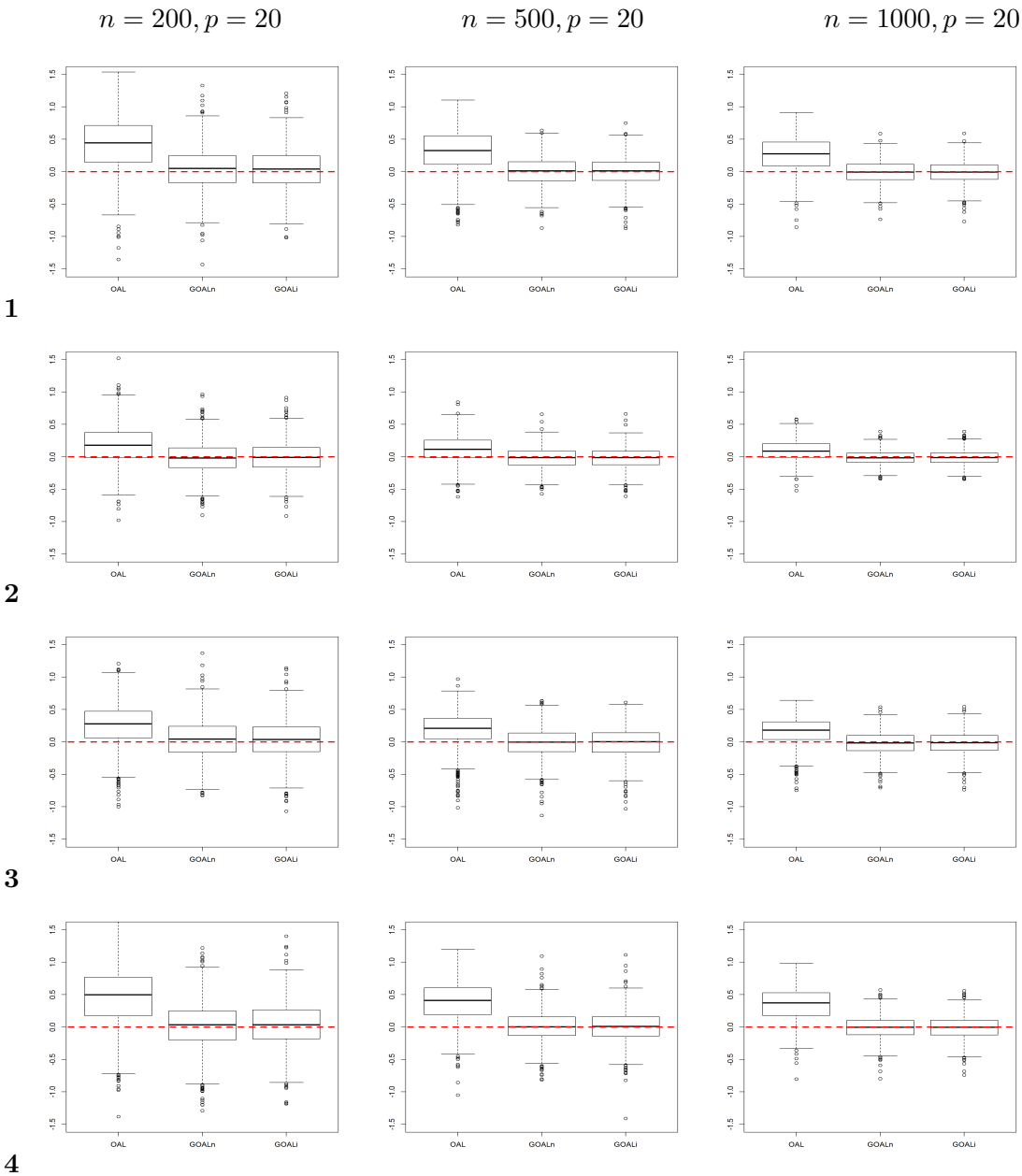


Figure 2.10 Box plots of 1000 IPTW estimates for the average treatment effect (ATE) for OAL, naive GOAL (GOALn) and penalized iteratively re-weighted least squares GOAL (GOALi) under Scenarios 1, 2, 3 and 4 (by row) in the low-dimensional settings with $\rho = 0.75$. The true value of ATE is indicated with dotted line (ATE=0).

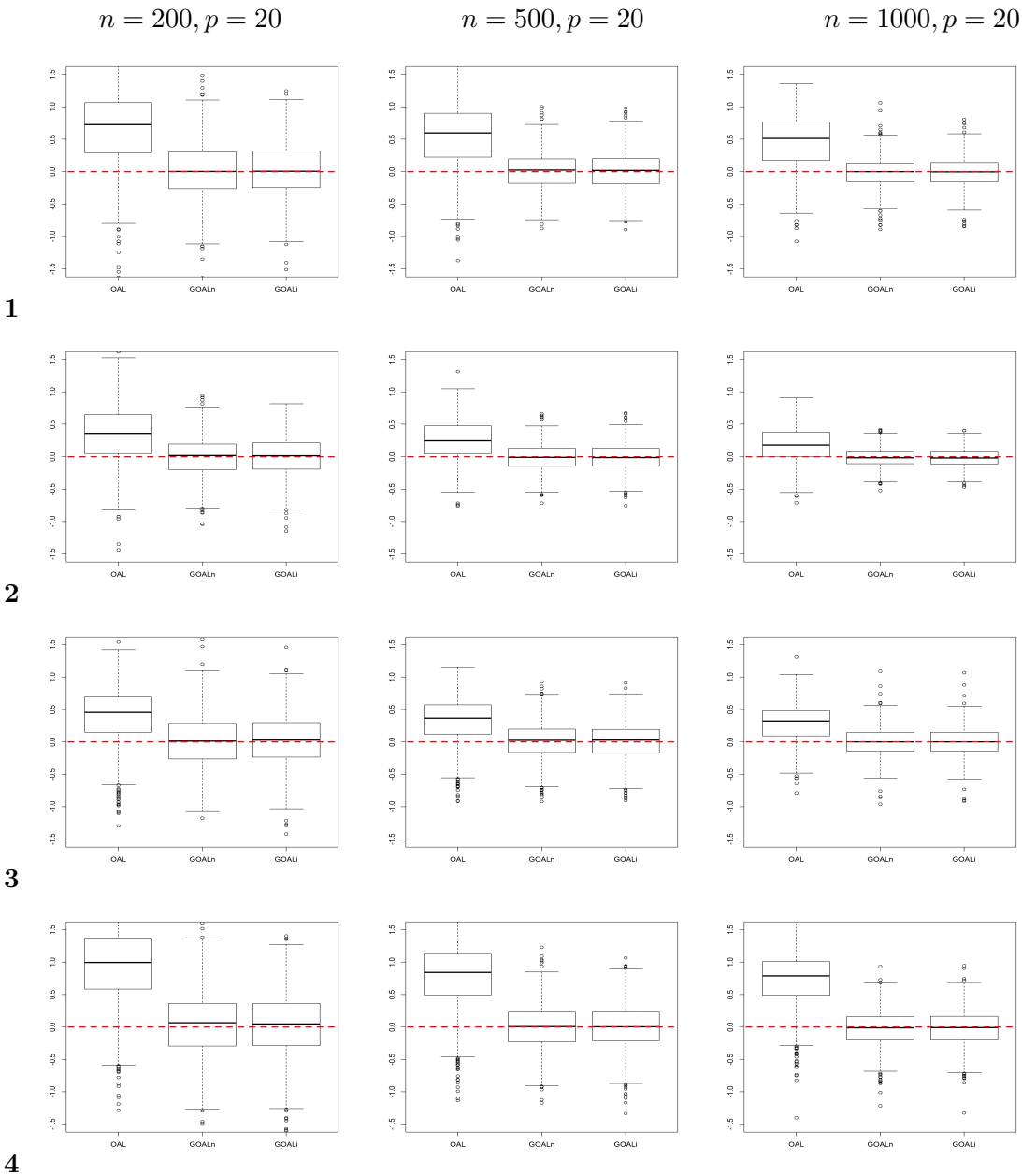


Figure 2.11 Weighted absolute mean difference (wAMD) between the exposure groups for OAL, GOALn and GOALi over 1000 simulations with $n = 200, p = 100$ and $\rho = 0$ under Scenarios 1, 2, 3 and 4 (by row).

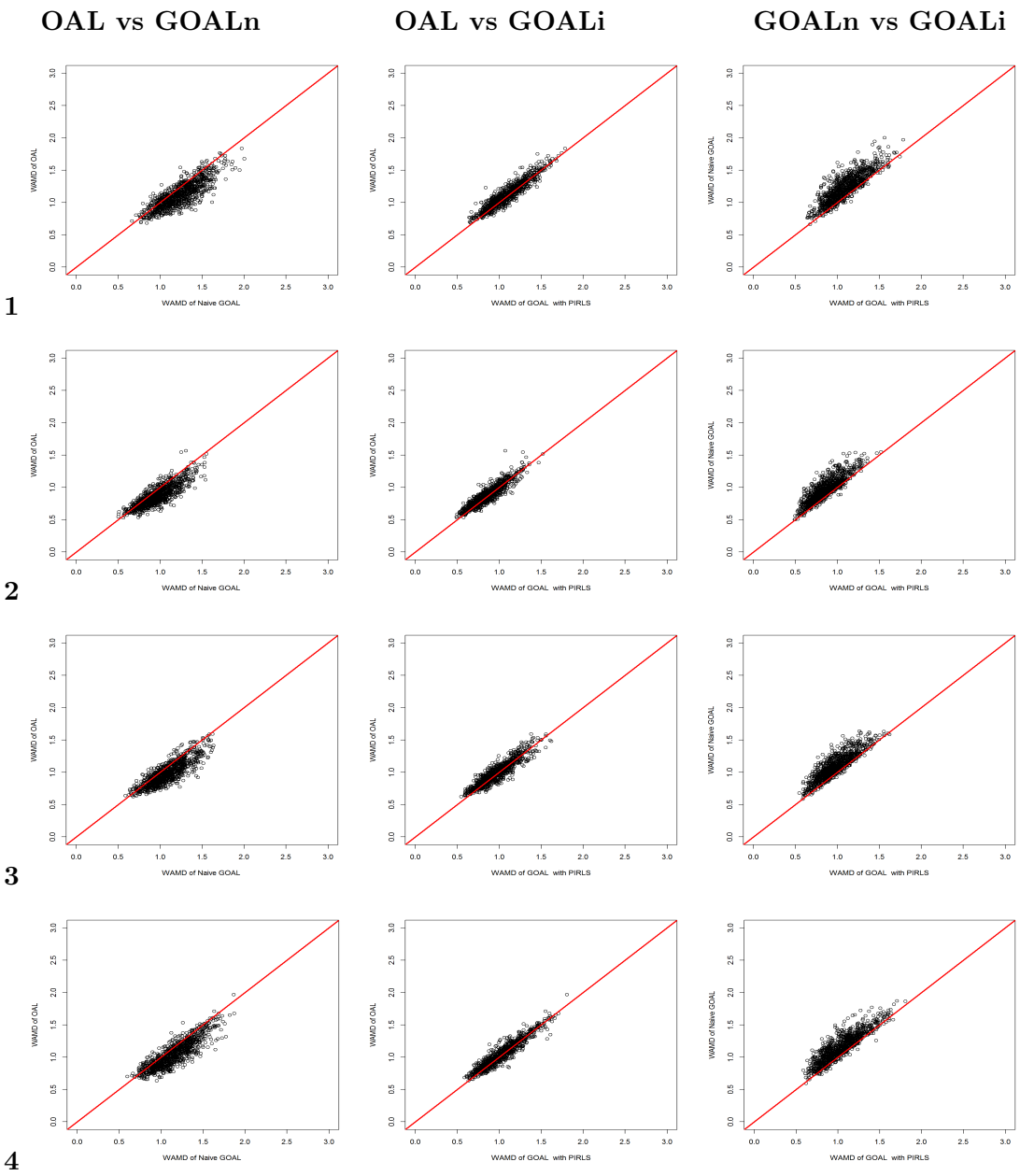


Figure 2.12 Weighted absolute mean difference (wAMD) between the exposure groups for OAL, GOALn and GOALi over 1000 simulations with $n = 200, p = 100$ and $\rho = 0.75$ under Scenarios 1, 2, 3 and 4 (by row).

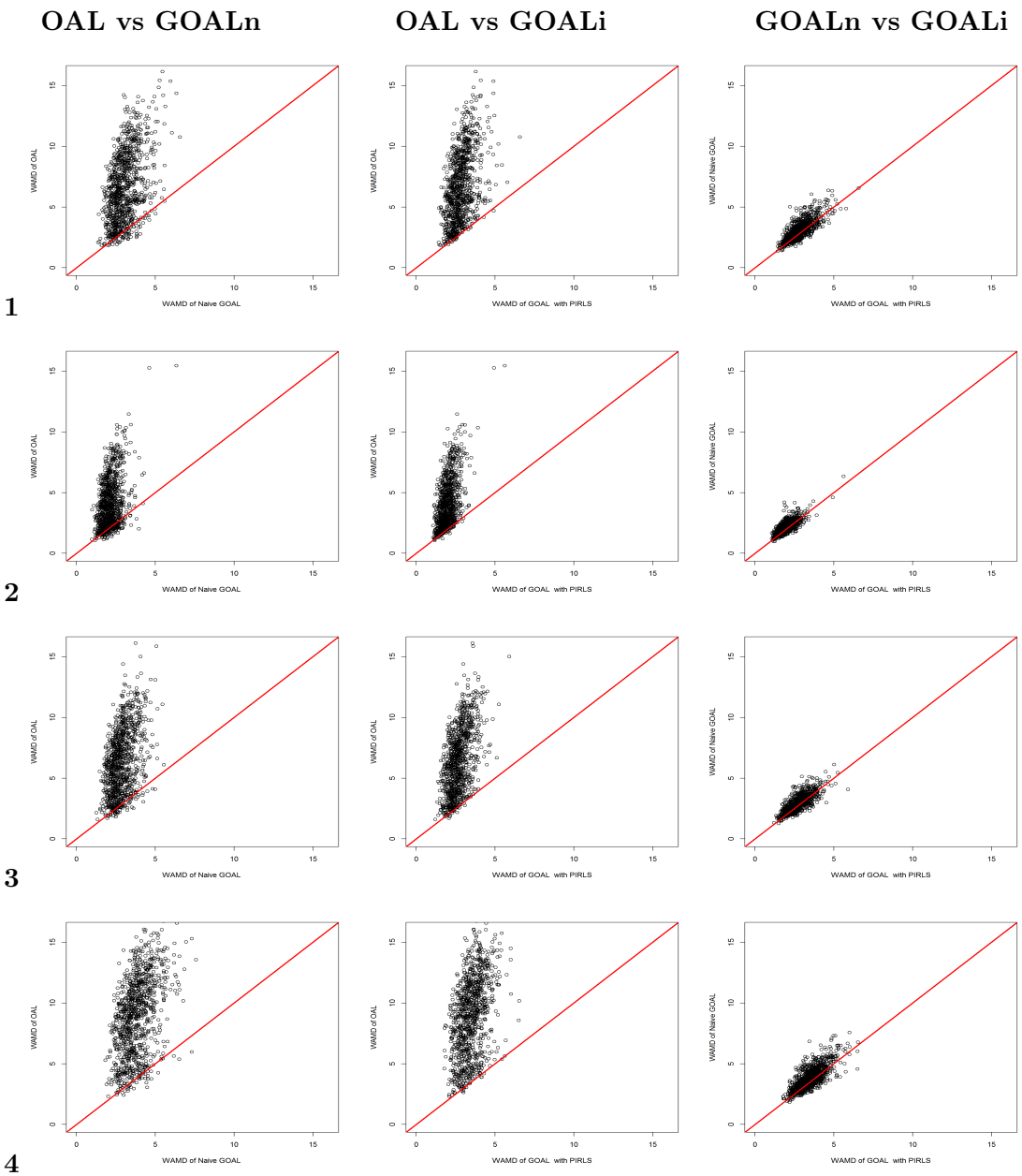


Figure 2.13 Weighted absolute mean difference (wAMD) between the exposure groups for OAL, GOALn and GOALi over 1000 simulations with $n = 200, p = 20$ and $\rho = 0$ under Scenarios 1, 2, 3 and 4 (by row).

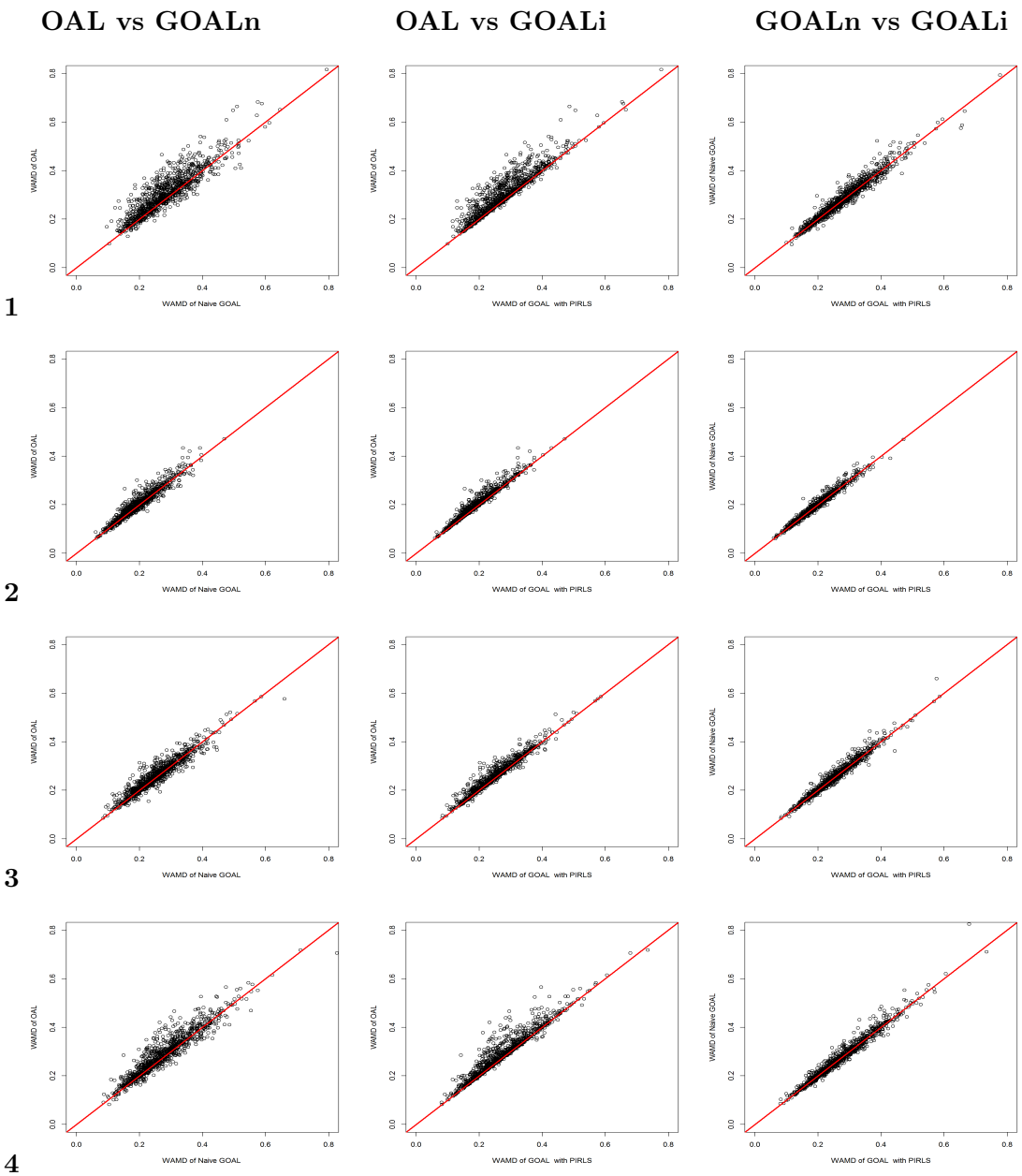


Figure 2.14 Weighted absolute mean difference (wAMD) between the exposure groups for OAL, GOALn and GOALi over 1000 simulations with $n = 200, p = 20$ and $\rho = 0.75$ under Scenarios 1, 2, 3 and 4 (by row).

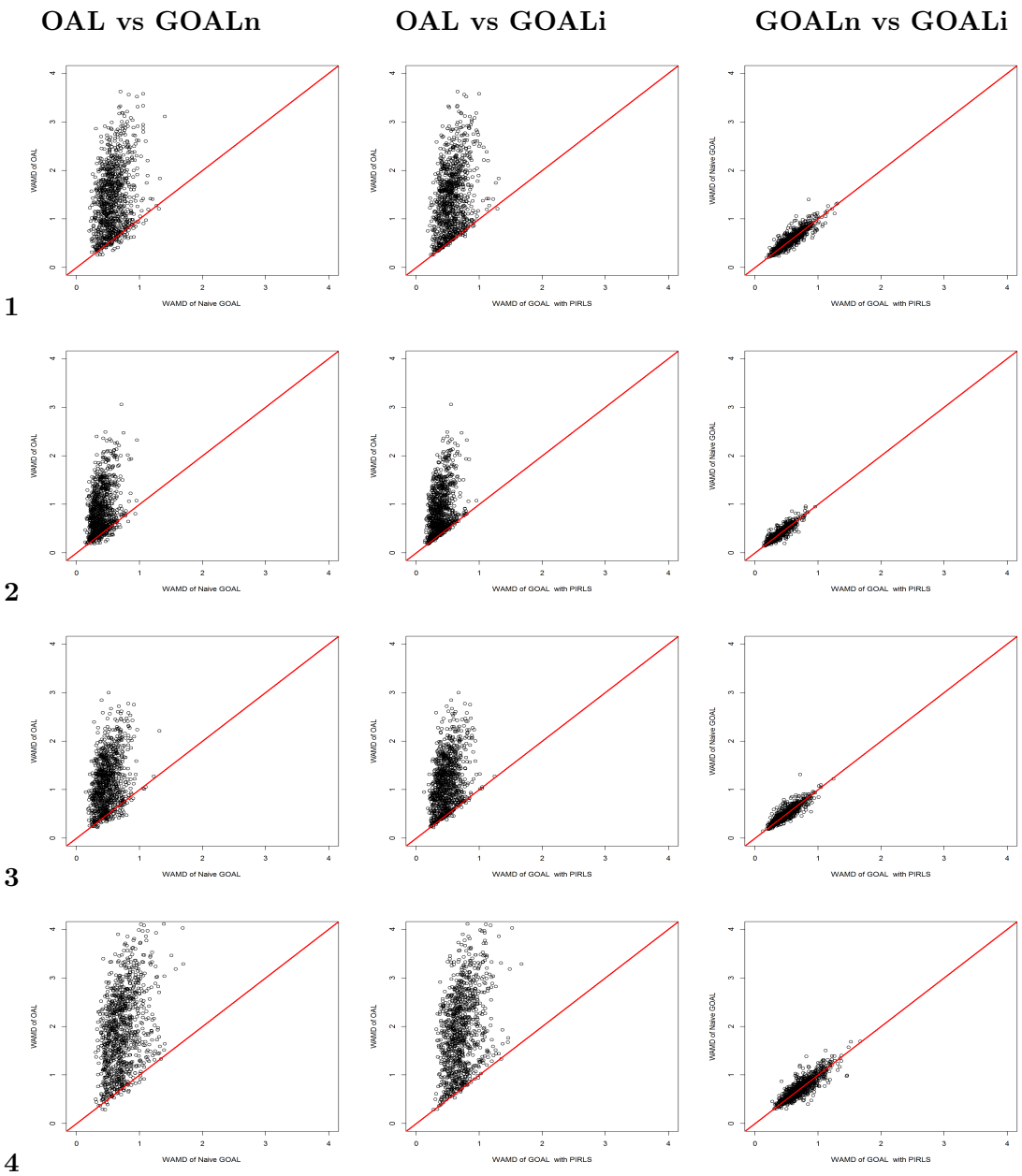


Figure 2.15 Probability of covariate being included in the propensity score (PS) model for OAL, naive GOAL (GOALn) and penalized iteratively re-weighted least squares GOAL (GOALi) under Scenarios 1, 2, 3 and 4 (by row) with $n = 200$ and $p = 100$.

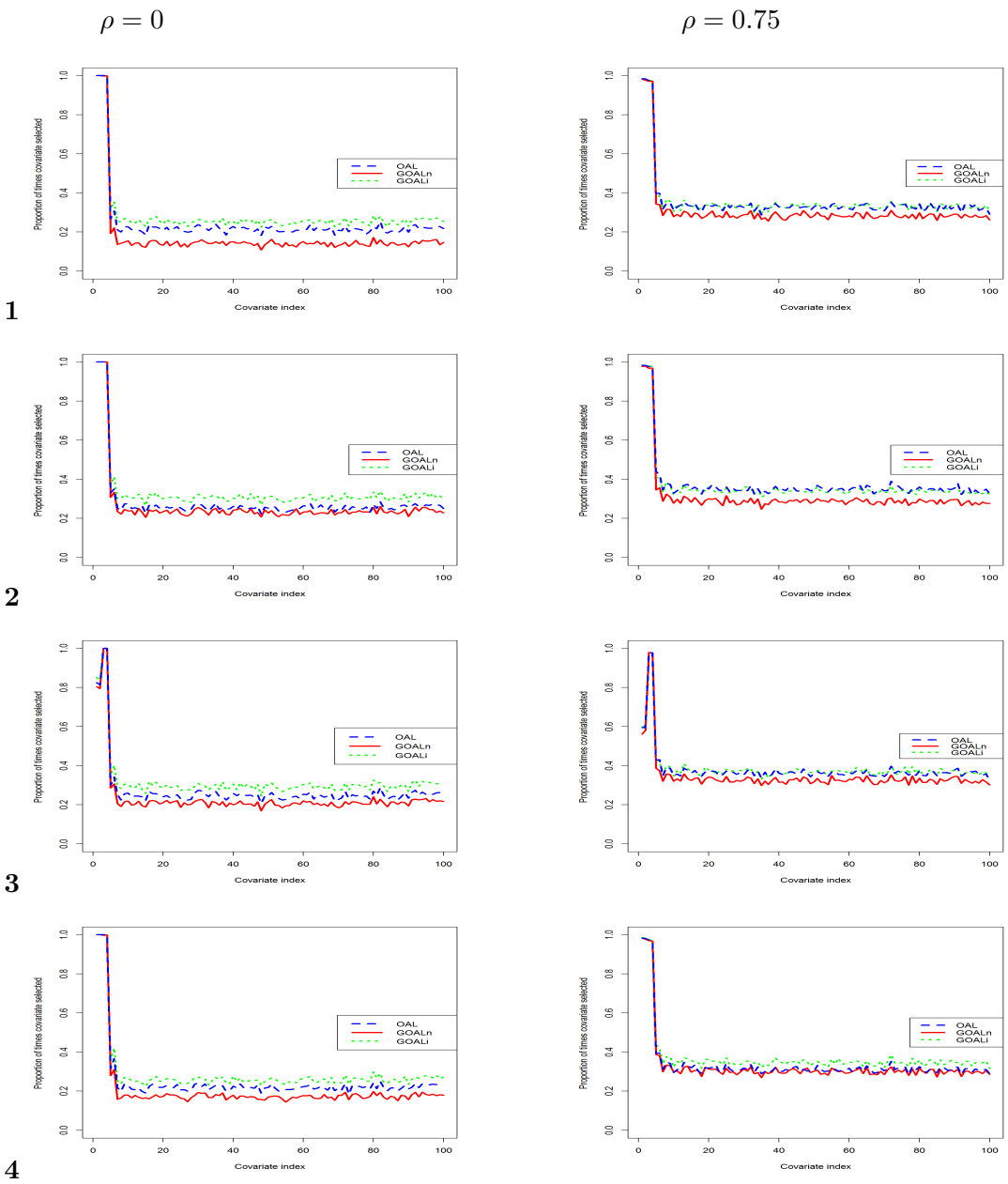
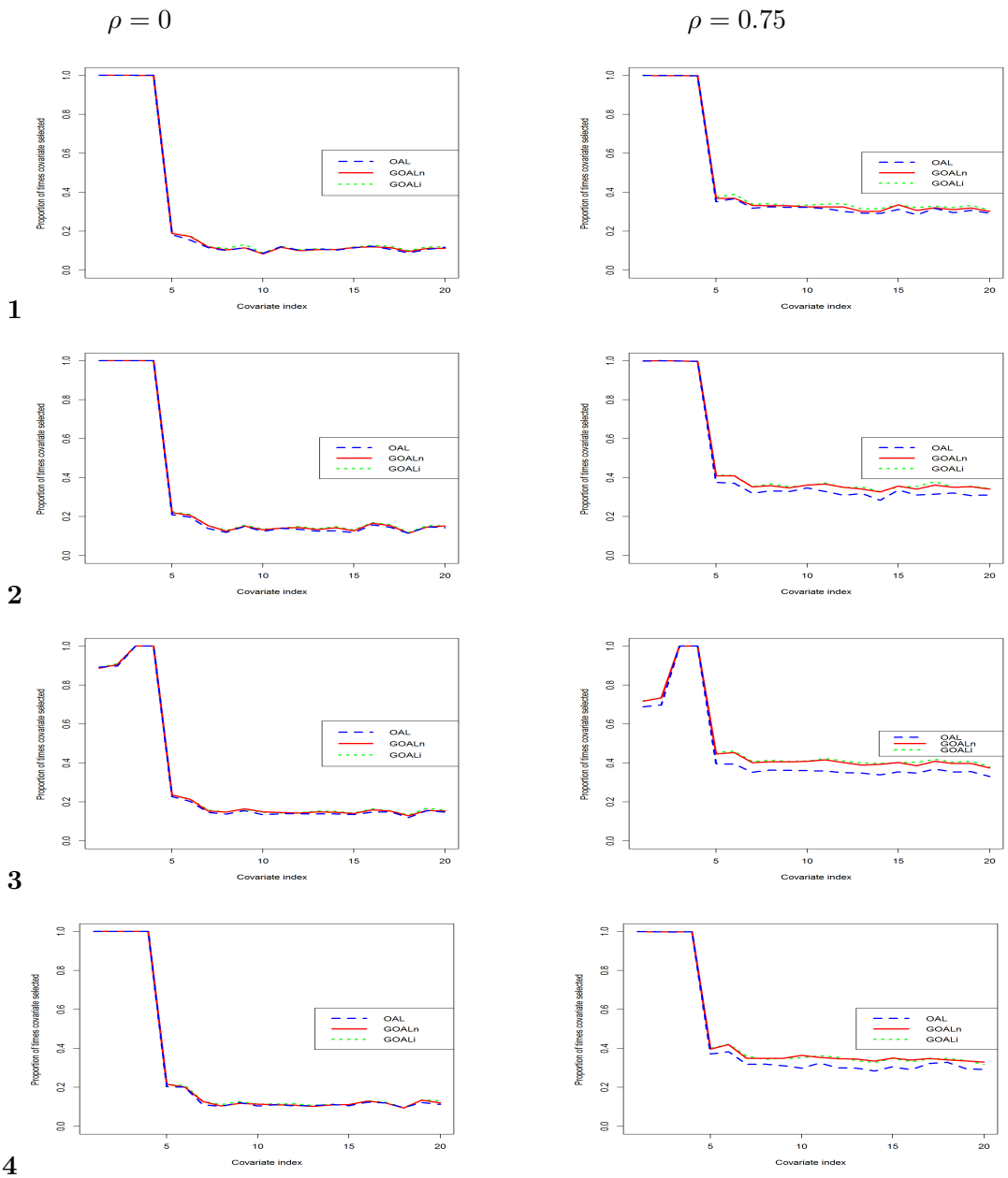


Figure 2.16 Probability of covariate being included in the propensity score (PS) model for OAL, naive GOAL (GOALn) and penalized iteratively re-weighted least squares GOAL (GOALi) under Scenarios 1, 2, 3 and 4 (by row) with $n = 200$ and $p = 20$.



CHAPITRE III

DEUXIÈME ARTICLE: OUTCOME-ADAPTIVE REGULARIZATION FOR CAUSAL MEDIATION ANALYSIS: VARIABLE SELECTION FOR DIRECT EFFECT AND INDIRECT EFFECT

Ismaila Baldé and Geneviève Lefebvre

Ce Chapitre 3 est constitué d'un article original dont je suis le premier auteur. Cet article est sur le point d'être soumis au journal *Statistics in Medicine*. Dans cet article, j'ai collaboré avec ma directrice de thèse, Geneviève Lefebvre, professeure au Département de mathématiques à l'UQAM. J'ai développé l'idée originale, effectué la revue de la littérature, réalisé les analyses statistiques et rédigé le code informatique et le manuscrit.

Outcome-adaptive regularization for causal mediation analysis: Variable selection for direct effect and indirect effect

Ismaila Baldé and Geneviève Lefebvre

Abstract: Propensity score (PS) methods are widely used when seeking to estimate the effect of a treatment (exposure) on an outcome from observational data. In the causal inference literature, it is nowadays well established that adjusting for confounders is necessary to obtain unbiased estimators of the treatment (exposure) effect and that adjusting for predictors of the outcome generally improves their efficiency. While variable selection for treatment effect estimation has attracted considerable attention in the past two decades, variable selection in causal mediation analysis is still an emerging topic. In this paper, we introduce a novel outcome-adaptive regularization approach for causal mediation analysis (MOAR) that aims to select desirable variables for the estimation of the treatment's direct and indirect effects, in presence or absence of intermediate confounders. Our MOAR approach relies on two adaptive regularization methods recently introduced for treatment effect estimation, namely the outcome-adaptive lasso (OAL) and the generalized outcome-adaptive lasso (GOAL). More precisely, our proposed approach uses (G)OAL for estimating the total effect and uses (G)OAL again on a residualized outcome for estimating the direct effect, while the indirect effect is deduced by taking the difference. Simulation results show that for the direct effect, MOAR selects the PS model that includes the confounders of the treatment-outcome and mediator-outcome relationships and the predictors of the outcome, while excluding the remaining variables. Variable selection using the MOAR approach is illustrated using data from the Harvard School of Public Health College Alcohol Study, 1999.

Keywords: Adaptive lasso, adaptive elastic net, causal mediation analysis, propensity score, variable selection.

3.1 Introduction

Propensity score (PS) methods are often used to control for confounding bias in observational studies (Rosenbaum et Rubin, 1983; Brookhart *et al.*, 2006). These methods require the specification of a model for the conditional probability of the treatment (or exposure) given a set of pre-treatment covariates. This set must include all confounders to achieve unbiased estimation of the effect of the treatment on the outcome, but other types of covariates may improve or worsen the performance of the treatment effect estimators. Rubin (1997) indicated that including covariates that are unrelated to the outcome (spurious covariates or pure predictors of treatment) in a PS model can lead to efficiency losses. Later, Patrick et al. (2011) found that the inclusion of covariates related only to the treatment inflates the standard error of PS treatment effect estimators. Lunceford and Davidian (2004) and Brookhart et al. (2006) showed that including pure predictors of the outcome in a PS model increases precision without increasing bias. Building on this knowledge, Shortreed and Ertefaie (2017) recently proposed the outcome-adaptive lasso (OAL), which selects the PS estimator that includes all confounders and outcome predictors, while excluding remaining covariates. Their approach is based on a regularized logistic PS model for the treatment, where the coefficients of this model are penalized inversely proportional to the association of the respective covariates with the outcome in a separate outcome model, and where the tuning parameter is selected by minimizing the weighted absolute mean difference (wAMD) between the treated and untreated groups. Koch et al. (2018) proposed an approach using group lasso and doubly robust estimation (GLiDeR) to identify confounders and predictors of outcome, and perform treatment effect estimation. Baldé et al. (2022) generalized OAL by using the elastic net penalty instead of the lasso to better handle high- and low-dimensional settings with correlated covariates.

In practice, researchers are interested in estimating the effect of the treatment on the outcome, but also often in understanding the mechanisms by which the treatment influences the outcome, which can be accomplished via mediation analysis (Valeri et VanderWeele,

2013). Mediation analysis is a statistical technique that relies on the idea that there exists a mediating variable (mediator) that is an intermediary step in the causal pathway between the treatment and outcome. The goal is then to estimate the magnitude of 1) the effect of treatment on the outcome through the mediator (indirect effect) and 2) the effect of the treatment on the outcome not due to the mediator (direct effect). Mediation analyses are increasingly performed in the fields of psychological, social and biomedical sciences (Valeri et VanderWeele, 2013; Nguyen *et al.*, 2021). Several estimation approaches have been proposed in the traditional and causal settings for estimating direct and indirect effects (Valeri et VanderWeele, 2013; Lange *et al.*, 2012; Imai *et al.*, 2010a). In the causal framework, the choice of covariates to include in the mediation model is crucial for achieving unbiased estimation of these effects. In particular, an expanded set of confounders is required to achieve unbiasedness (Valeri et VanderWeele, 2013). It is therefore relevant to ask how the knowledge and algorithms developed for the selection of variables for the estimation of the total effect of treatment is transposed in the context of causal mediation analysis.

To our knowledge, only few works have addressed confounder selection in the causal mediation literature (Diop *et al.*, 2021; Jones *et al.*, 2021; Farbmacher *et al.*, 2021). Diop et al. (2021) considered both a model-based regression approach (Imai *et al.*, 2010a) and an inverse-probability-weighting approach (Lange *et al.*, 2012) for mediation, and investigated the impact of adjusting for pure predictors of the treatment, mediator, and outcome on the standard error of the corresponding natural direct and indirect effects estimators. For both approaches they found that adjusting for pure predictors of the treatment results in an efficiency loss for the natural direct and indirect effects estimators, while adjusting for pure predictors of the outcome leads to an efficiency gain for these estimators. They also found that adjusting for pure predictors of the mediator increases the standard error of the natural direct effect estimator, while the impact on the standard error of the natural indirect effect estimator can be either positive or negative, depending on the effect of the mediator on the outcome and the effect of the treatment on the mediator (Diop *et al.*, 2021). Ye et al. (2021) also targeted the

inverse-probability-weighting approach of Lange et al. (2012), but proposed using the OAL approach and more standard lasso regularization approaches for the estimation of the PS models for a binary treatment and a binary mediator. In their work based on OAL, the tuning parameters were either selected by minimizing the wAMD or the deviance. As in Diop et al. (2021), they found that adjusting for covariates that are only related to the outcome can significantly increase the efficiency of the natural direct and indirect effects estimators (Ye *et al.*, 2021).

It is frequent that prior knowledge regarding the type of covariates (that is, confounder, predictor of treatment, predictor of outcome, or predictor of mediator) used in a mediation analysis is incomplete, thereby limiting the applicability of the findings by Diop et al. (2021) in practice. Ye et al. (2021) proposed a data-driven algorithm for the selection of the treatment and mediator PS models, which can be helpful when the knowledge regarding data structure is incomplete. While their approach is interesting, some questions remain regarding its functioning and empirical performance. In particular, the adaptive weights used in the application of OAL are the same for both PS models (treatment and mediator), and are specified using an outcome model that conditions on the covariates, treatment and mediator. This strategy departs from the original OAL approach wherein the adaptation is built from an outcome model which excludes potential mediating variables. While Ye et al. (2021)'s adaptive weights building strategy does not appear unreasonable at first sight, it is hypothesized it could cause OAL to exclude covariates causing the outcome indirectly through the mediator (such as treatment-mediator confounders). Second, although the authors did investigate which sets of covariates should be included in the PS models for calculating the weights, the insight is only partial as the simulation setting only considered one confounding variable for the treatment, mediator and outcome altogether. This is particularly a concern given above remark on the possible exclusion of treatment-mediator confounders. Lastly, although it is possible to use weighting ideas with a continuous treatment or a continuous mediator, the OAL method is not readily applicable for the estimation of conditional densities, which restricts the application of the approach by Ye et al. (2021) in these situations.

In this paper, we aim to expand knowledge about variable selection for unbiased and efficient direct and indirect effects estimation and propose a data-driven approach based on regularized regression to accomplish this. While Ye et al. (2021) apply OAL to select the treatment PS model and the mediator PS model in a separate fashion, we instead propose a two-step (sequential) application of OAL to select appropriate variables for estimating the direct and indirect effects of the treatment on the outcome, approach hereafter named MOAL. More precisely, our proposed approach uses OAL for estimating the total effect and uses OAL again on a residualized outcome for estimating the direct effect, while the indirect effect is deduced by taking the difference. As in Ye et al. (2021), our approach based on OAL considers a binary treatment, but the mediator can be either continuous or binary. Moreover, as the generalized OAL (GOAL) (Baldé *et al.*, 2022) was found a competitive alternative to OAL in the context of total treatment effect estimation, we also investigate the performance of GOAL in the proposed variable selection approach for mediation (MGOAL). In addition, we extend MOAL and MGOAL approaches to accommodate the presence of intermediate confounders, which are common in causal mediation contexts (Acharya *et al.*, 2016; Imai *et al.*, 2010b). In the sequel, MOAL and MGOAL are referred under the umbrella term MOAR, which stands for Mediation Outcome Adaptive Regularization.

The rest of the paper is organized as follows. In Section 2, we introduce the proposed MOAR approach. We briefly recall the OAL approach in Section 2.1 and the GOAL approach in Section 2.2. We then provide two simple algorithms for the respective implementation of MOAL and MGOAL in Section 2.3. These algorithms are introduced for a continuous outcome, a continuous mediator and a binary treatment. In Section 2.4, we explain how MOAR can be modified to deal with intermediate confounding, which yields the approach named MOARI. In Section 3, we describe our simulation study and analyze the results to assess the empirical properties of MOAR and MOARI. In Section 4, we apply MOAR on real data, in order to evaluate the effect of marijuana use on grade point average, with school disengagement considered as a potential mediator. We conclude the paper with a discussion in Section 5.

3.2 Methods

In the sequel, we consider a continuous outcome Y, a continuous mediator M, a binary treatment A ($A = 1$: treated and $A = 0$: untreated) and a baseline covariate vector $X = (X_1, \ldots, X_p)$. Therefore, the observed data can be written as (X_i, A_i, M_i, Y_i), where i indexes the n independent observations ($i = 1, \ldots, n$). We assume that the covariates X are measured prior to the treatment A, which is measured prior to the mediator M, which in turn is measured prior to the outcome Y. In this article, we focus on estimating the average total effect (ATE) and the average controlled direct effect (ACDE). Under the assumption of no treatment-mediator interaction, which we consider herein, the natural direct effect is equivalent to the controlled direct effect and the difference between the total effect and the direct effect can be interpreted as an indirect effect (Valeri et VanderWeele, 2013; Vansteelandt, 2009). The average total effect is defined as ATE=$E[Y^{a=1} - Y^{a=0}]$, where $Y^{a=1}$ denotes the counterfactual outcome under treatment and $Y^{a=0}$ the counterfactual outcome under no treatment. The average controlled direct effect is defined as ACDE=$E[Y^{a=1,m} - Y^{a=0,m}]$, where $Y^{a=j,m}$ denotes the counterfactual outcome when the treatment is set to j ($j = 0$ or $j = 1$) and the mediator is set to m. Then, the average indirect effect (AIE) can be written as: AIE=ATE-ACDE.

Causal inference from observational studies is based on several assumptions. The following four assumptions are required to guarantee unbiased estimators of the ATE: positivity, consistency, exchangeability and stable unit treatment value assumption (Shortreed et Ertefaie, 2017). Positivity can be written as $0 < P(A = 1|X = x) < 1$ for all possible x values. It means that the probability of receiving both levels of treatment conditional on X is positive for all individuals. Consistency is defined as $Y = AY^{a=1} + (1 - A)Y^{a=0}$. That is, the observed outcome for an individual is equal to the counterfactual outcome under the treatment assignment the individual actually received. Exchangeability or no unmeasured confounding assumptions means X includes all possible confounders: Y^a II $A \mid X$. Stable unit treatment value assumption means each individual's coun-

terfactual outcomes are not influenced by the treatment status of other individuals: $(Y_i^{a=0}, Y_i^{a=1}) \amalg A_j$, for $i \neq j$ (Zhu, 2021). he average natural direct effect (ANDE) and average natural indirect effect (ANIE) can be identified if the following four assumptions are satisfied:

$\mathcal{A}_1$: $Y_{am} \amalg A \mid X$. That is, there are no unmeasured treatment-outcome confounders given X.

$\mathcal{A}_2$: $Y_{am} \amalg M \mid X, A$. That is, there are no unmeasured mediator-outcome confounders given (X, A).

$\mathcal{A}_3$: $M_a \amalg A \mid X$. That is, there are no unmeasured treatment-mediator confounders given X.

$\mathcal{A}_4$: $Y_{am} \amalg M_{a^*} \mid X$. This assumption does not hold when there are (measured or unmeasured) mediator-outcome confounders affected by treatment (Lindmark, 2021).

In addition, positivity, consistency, and composition assumptions are required for ANDE and ANIE. Moreover, also note that if the first two assumptions ($\mathcal{A}_1$ and $\mathcal{A}_2$) as well as positivity and consistency hold, then the average controlled direct effect (ACDE) can be identified (VanderWeele, 2018). The presence of intermediate confounders (Z) (mediator-outcome confounders affected by treatment) causes problems in the identification of the natural effects as well as the controlled direct effect. While the ANDE and the ANIE cannot be identified in the presence of Z without strong individual-level homogeneity assumptions (Acharya *et al.*, 2016), the ACDE is nonparametrically identified under sequential unconfoundedness assumptions (Robins, 1997; Acharya *et al.*, 2016). The sequential unconfoundedness assumptions include the assumption $\mathcal{A}_1$ and additional assumptions $\mathcal{A}_5$ and $\mathcal{A}_6$, which are: $\mathcal{A}_5$: $Y_{am} \amalg M \mid X, A, Z$ and $\mathcal{A}_6 : 0 < P(M = m|A = 1, X = x, Z = z) < 1$. The assumption $\mathcal{A}_5$ means that there are no unmeasured confounders (affected by treatment or not) for the effect of the mediator on the outcome, conditional on the treatment (A), pretreatment confounders (X), and intermediate confounders (Z) (Acharya *et al.*, 2016). Assumption $\mathcal{A}_6$ means all

mediator values conditional on the treatment (A), pretreatment confounders (X), and intermediate confounders (Z) have non-zero probability (Lange *et al.*, 2017).

We start by recalling the OAL (Shortreed et Ertefaie, 2017) and GOAL (Baldé *et al.*, 2022) approaches. We then present the two versions of our proposed MOAR approach, namely MOAL and MGOAL.

3.2.1 OAL Approach [Shortreed and Ertefaie, 2017]

We present a short review of the OAL aproach proposed by Shortreed and Ertefaie (2017), which is devised to estimate the ATE. We assume the following PS model parametrized by α:

$$logit\left\{\pi(X,\alpha)\right\} = logit\left\{P(A=1|X)\right\} = \sum_{j=1}^{p}\alpha_j X_j. \tag{3.1}$$

Shortreed and Ertefaie (2017) use the adaptive lasso penalty to estimate the ideal PS model, which includes all confounders ($X_{\mathcal{C}}$) to avoid bias and all predictors of the outcome ($X_{\mathcal{P}}$) to increase statistical efficiency, while excluding the predictors of the treatment ($X_{\mathcal{I}}$) and spurious variables ($X_{\mathcal{S}}$):

$$logit\left\{\pi(X,\hat{\alpha})\right\} = logit\left\{\widehat{P}(A=1|X)\right\} = \sum_{j\in\mathcal{C}}\hat{\alpha}_j X_j + \sum_{j\in\mathcal{P}}\hat{\alpha}_j X_j. \tag{3.2}$$

The OAL estimator is defined as:

$$\hat{\alpha}(OAL) = \arg\min_{\alpha}\left[\sum_{i=1}^{n}\left\{-a_i(x_i^T\alpha) + \log\left(1+e^{x_i^T\alpha}\right)\right\} + \lambda\sum_{j=1}^{p}\hat{w}_j|\alpha_j|\right], \tag{3.3}$$

where $\lambda > 0$, $\hat{w}_j = \left|\hat{\beta}_j^{ols}\right|^{-\gamma}$ such that $\gamma > 1$ and $(\hat{\beta}_A^{ols}, \hat{\beta}^{ols}) = \arg\min_{(\beta_A,\beta)}\|Y - \beta_A A - \mathbf{X}\beta\|_2^2$.

Shortreed and Ertefaie (2017) used the fitted PS model from (3.3) to construct the following inverse probability of treatment (IPT) weights:

$$\hat{\tau}_i^{\lambda} = \frac{A_i}{\pi_i^{\lambda}\left\{X_i,\hat{\alpha}(OAL)\right\}} + \frac{1-A_i}{1-\pi_i^{\lambda}\left\{X_i,\hat{\alpha}(OAL)\right\}},$$

where $\lambda \in S_\lambda = \{n^{-10}, n^{-5}, n^{-2}, n^{-1}, n^{-0.75}, n^{-0.5}, n^{-0.25}, n^{0.25}, n^{0.49}\}$. The optimal λ in the set S_λ (referred to herein as $\hat{\lambda}$) minimizes the weighted absolute mean difference (wAMD) between the treated and untreated groups (Shortreed et Ertefaie, 2017):

$$wAMD(\lambda) = \sum_{j=1}^{p} \left|\hat{\beta}_j^{ols}\right| \left| \frac{\sum_{i=1}^{n} \hat{\tau}_i^{\lambda} X_{ij} A_i}{\sum_{i=1}^{n} \hat{\tau}_i^{\lambda} A_i} - \frac{\sum_{i=1}^{n} \hat{\tau}_i^{\lambda} X_{ij}(1-A_i)}{\sum_{i=1}^{n} \hat{\tau}_i^{\lambda}(1-A_i)} \right|. \tag{3.4}$$

The estimated average total effect ($\widehat{ATE}$) is based on the optimal $\hat{\lambda}$ and is defined as

$$\widehat{ATE} = \frac{\sum_{i=1}^{n} \hat{\tau}_i^{\hat{\lambda}} Y_i A_i}{\sum_{i=1}^{n} \hat{\tau}_i^{\hat{\lambda}} A_i} - \frac{\sum_{i=1}^{n} \hat{\tau}_i^{\hat{\lambda}} Y_i (1-A_i)}{\sum_{i=1}^{n} \hat{\tau}_i^{\hat{\lambda}} (1-A_i)}.$$

Shortreed and Ertefaie (2017) provided `R` code to implement their method, which can be found in the Supplementary Material of their article (https://onlinelibrary.wiley.com/doi/10.1111/biom.12679).

3.2.2 GOAL Approach [Baldé et al., 2022]

Baldé et al. (2022) assume the same PS model as in Shortreed and Ertefaie (2017) and consider an adaptive elastic net regularization instead of the lasso to define GOAL. More precisely, the GOAL approach starts from an outcome-adaptive elastic net problem which is then transformed into Shortreed and Ertefaie (2017)'s outcome-adaptive lasso representation by augmenting the data suitably. The adaptive elastic net regularization underlying GOAL handles problems arising from high-dimension and collinearity simultaneously.

The GOAL problem is defined as

$$\hat{\alpha}(GOAL) = \arg\min_{\alpha} \left[\sum_{i=1}^{n} \left\{ -a_i(x_i^T \alpha) + \log\left(1 + e^{x_i^T \alpha}\right) \right\} + \lambda_1 \sum_{j=1}^{p} \hat{w}_j |\alpha_j| + \lambda_2 \sum_{j=1}^{p} \alpha_j^2 \right], \tag{3.5}$$

where $\lambda_1, \lambda_2 > 0$ and $\hat{w}_j$ is defined as in (3.3); $S_{\lambda_1} = S_\lambda$ and
$S_{\lambda_2} = \{0, 10^{-2}, 10^{-1.5}, 10^{-1}, 10^{-0.75}, 10^{-0.5}, 10^{-0.25}, 10^{0}, 10^{0.25}, 10^{0.5}, 10^{1}\}$.
In Baldé et al. (2022) two versions of GOAL were introduced (naive and iterative) to solve problem (3.5), where these differ by the way the data augmentation is implemented.

We present below the naive GOAL algorithm.

GOAL algorithm

1. Input: Design matrix and treatment data $(\mathbf{X}, A)$ and S_{λ_2};
2. For a fixed $\lambda_2 \in S_{\lambda_2}$ define: $\mathbf{X}^* = \begin{pmatrix} \mathbf{X} \\ \sqrt{\lambda_2}\mathbf{I}_p \end{pmatrix}$ and $A^* = \begin{pmatrix} A \\ 0_p \end{pmatrix}$;
3. Estimate the PS model with OAL on augmented data $(\mathbf{X}^*, A^*)$, that is solve
$$\hat{\alpha}_N^*(\text{naive}) = \arg\min_{\alpha} \left[\sum_{i=1}^{n+p} \left\{ -a_i^*(x_i^{*T}\alpha) + \log\left(1 + e^{x_i^{*T}\alpha}\right) \right\} + \lambda_1 \sum_{j=1}^{p} \hat{w}_j |\alpha_j| \right];$$
4. Compute $\hat{\alpha}_N(\text{adaptive elastic net}) = (1 + \lambda_2)\hat{\alpha}_N^*(\text{naive})$;
5. Output: $\hat{\alpha}_N(\text{adaptive elastic net})$.

To select the tuning parameters (λ_1, λ_2), Baldé et al. (2022) balance the exposure groups on a two-dimensional surface and choose the parameters corresponding to the smallest wAMD. We refer the reader to Baldé et al. (2022) for details and `R` code to implement GOAL.

3.2.3 The proposed MOAR Approach

The assumptions required for the causal interpretation of the direct and indirect effects can be usefully characterized with a causal mediation diagram (Valeri et VanderWeele, 2013). Consider the directed acyclic graph (DAG) in Figure 3.1, displaying the posited causal relationships between the variables.

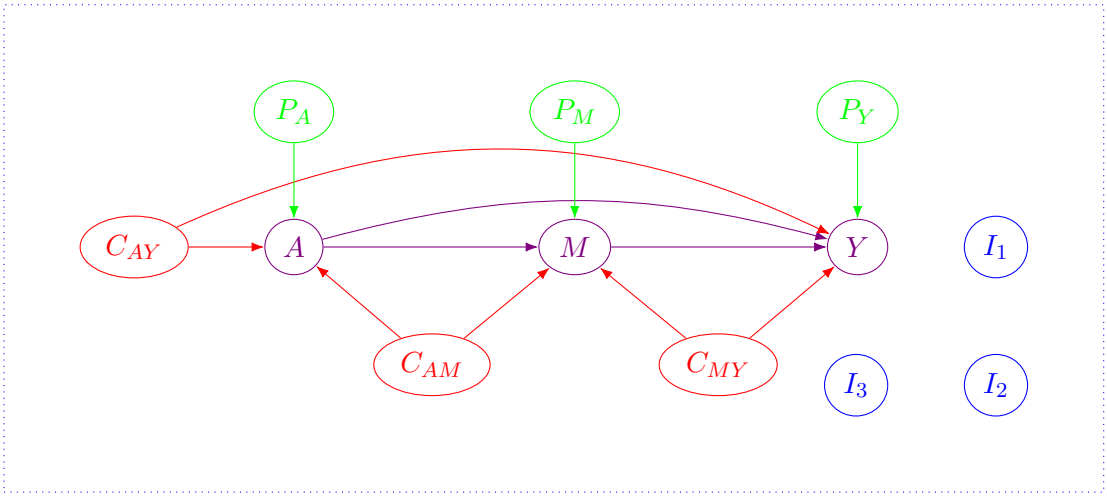


Figure 3.1 Hypothesized causal mediation graph.

The variables represented in the DAG are the covariates $X = (C_{AY}, C_{MY}, C_{AM}, P_Y, P_A, P_M, I_S)$, treatment (A), mediator (M) and outcome (Y). The baseline covariates (X) include treatment-outcome confounders (C_{AY}), mediator-outcome confounders (C_{MY}), treatment-mediator confounders (C_{AM}), pure predictors of outcome (P_Y), pure predictors of treatment (P_A), pure predictors of mediator (P_M) and spurious covariates $(I_S = (I_1, I_2, I_3))$.

Let $\mathscr{C}_{\mathscr{A}\mathscr{Y}}$, $\mathscr{C}_{\mathscr{M}\mathscr{Y}}$, $\mathscr{C}_{\mathscr{A}\mathscr{M}}$, $\mathscr{P}_{\mathscr{Y}}$, $\mathscr{P}_{\mathscr{M}}$, $\mathscr{P}_{\mathscr{A}}$ and $\mathscr{I}_{\mathscr{S}}$ denote the indices of C_{AY}, C_{MY}, C_{AM}, P_Y, P_M, P_A and I_S, respectively. We assume a logistic PS model parameterized by α

$$logit\left\{\pi(X,\alpha)\right\} = logit\left\{P(A=1|X)\right\} = \sum_{j=1}^{p} \alpha_j X_j.$$

Our objective is to use variable selection methods to obtain the following PS models:

$$logit\left\{\pi_{ATE}(X,\hat{\alpha})\right\} = \sum_{j\in\mathscr{C}_{\mathscr{A}\mathscr{Y}}} \hat{\alpha}_j X_j + \sum_{j\in\mathscr{C}_{\mathscr{A}\mathscr{M}}} \hat{\alpha}_j X_j + \sum_{j\in\mathscr{C}_{\mathscr{M}\mathscr{Y}}} \hat{\alpha}_j X_j + \sum_{j\in\mathscr{P}_{\mathscr{Y}}} \hat{\alpha}_j X_j + \sum_{j\in\mathscr{P}_{\mathscr{M}}} \hat{\alpha}_j X_j, \tag{3.6}$$

and

$$logit\left\{\pi_{ACDE}(X,\hat{\alpha})\right\} = \sum_{j\in\mathscr{C}_{\mathscr{A}\mathscr{Y}}} \hat{\alpha}_j X_j + \sum_{j\in\mathscr{C}_{\mathscr{M}\mathscr{Y}}} \hat{\alpha}_j X_j + \sum_{j\in\mathscr{P}_{\mathscr{Y}}} \hat{\alpha}_j X_j. \tag{3.7}$$

Equation (3.6) refers to the target PS model for the estimation of the ATE. Indeed, while the confounders C_{AM} refer specifically to the confounders of the treatment-mediator re-

lationship, these are also confounders of the treatment-outcome relationship when no conditioning on M is done. Therefore, data-driven confounder selection algorithms devised for standard (non-mediated) treatment effect estimation should keep these covariates. Using a similar reasoning, pure predictors of the mediator are also pure predictors of the outcome when no adjustment for M is done and should therefore be selected for efficiency purposes. Moreover, while the confounders for the mediator-outcome relationship are not required for the unbiased estimation of the ATE, they do predict the outcome hence their inclusion in the PS model of Equation (3.6). The Equation (3.7) refers to the target PS model for the estimation of the ACDE. It is useful to view this model as the target PS model for the estimation of the causal effect of the treatment on a residualized outcome in which the effect of the mediator on the outcome has been removed (that is, corresponding to removal of arrow $M \rightarrow Y$ in Figure 1). In this case, the confounders for the mediator-outcome relationship (C_{MY}) should be selected for the unbiased estimation of this treatment effect. However, the confounders of the treatment-mediator relationship (C_{AM}) and the pure predictors of the mediators (P_M) should not be selected as they do not cause the residualized outcome.

As a first stage, our proposed MOAR approach uses either OAL or GOAL to target the PS model in Equation (3.6) for estimating the ATE. A second stage considers the aforementioned residualized outcome and applies OAL or GOAL to target the PS model in Equation (3.7) for estimating the ACDE. We note that working with such a residualized outcome (that is, with the effect of M removed) for the estimation of the ACDE has been previously proposed by Vansteelandt (2009) (approach referred therein as sequential g-estimation, see Web Appendix A in Section 3.6.1). The AIE is then calculated as the difference between the estimated ATE and ACDE. We summarize the two stages to implement MOAR in the Algorithm's Part I and Part II, respectively. The decision to use either OAL or GOAL in both stages is made in step 4, leading to MOAL or MGOAL. In the presentation of the algorithm, the tuning parameters λ for OAL and λ_1 and λ_2 for GOAL are considered as fixed for simplicity. In practice, the selection of the tuning parameters is done separately in each stage by selecting the tuning parameter value (or values) that minimize the wAMD, as discussed in Section 3.2.1 (3.2.2).

Algorithm MOAR Part I: Estimation of average total effect (ATE)

1: Given original data $(\mathbf{X}, A, M, Y)$;

2: Compute $(\hat{\beta}_A^{ols}, \hat{\beta}^{ols}) = \arg\min_{(\beta_A,\beta)} \|Y - \beta_A A - \mathbf{X}\beta\|_2^2$;

3: Set $\hat{w}_j = \left|\hat{\beta}_j^{ols}\right|^{-\gamma}$ with $\gamma > 1$, $j = 1, 2, \ldots, p$;

4: Call

a: OAL algorithm to solve

$$\hat{\alpha} = \arg\min_{\alpha} \left[\sum_{i=1}^{n} \left\{ -a_i (x_i^T \alpha) + \log\left(1 + e^{x_i^T \alpha}\right) \right\} + \lambda \sum_{j=1}^{p} \hat{w}_j |\alpha_j| \right];$$

b: GOAL algorithm to solve

$$\hat{\alpha} = \arg\min_{\alpha} \left[\sum_{i=1}^{n+p} \left\{ -a_i^* ({x_i^*}^T \alpha) + \log\left(1 + e^{{x_i^*}^T \alpha}\right) \right\} + \lambda_1 \sum_{j=1}^{p} \hat{w}_j |\alpha_j| \right];$$

6: Compute predicted probabilities of exposure: $\hat{\pi}_i \{X_i, \hat{\alpha}\}$;

7: Compute IPT weights: $\hat{\tau}_i = \frac{A_i}{\hat{\pi}_i\{X_i,\hat{\alpha}\}} + \frac{1-A_i}{1-\hat{\pi}_i\{X_i,\hat{\alpha}\}}$;

8: Compute total effect: $\widehat{ATE} = \frac{\sum_{i=1}^{n} \hat{\tau}_i Y_i A_i}{\sum_{i=1}^{n} \hat{\tau}_i A_i} - \frac{\sum_{i=1}^{n} \hat{\tau}_i Y_i (1-A_i)}{\sum_{i=1}^{n} \hat{\tau}_i (1-A_i)}$.

Algorithm MOAR Part II: Estimation of average direct and indirect effects (ACDE and AIE)

1: Compute $(\tilde{\beta}_A^{ols}, \tilde{\beta}_M^{ols}, \tilde{\beta}^{ols}) = \arg\min_{(\beta_A,\beta_M,\beta)} \|Y - \beta_A A - \beta_M M - \mathbf{X}\beta\|_2^2$;

2: Define $Y^{res} = Y - \tilde{\beta}_M^{ols} M$;

3: Set $\tilde{w}_j = \left|\tilde{\beta}_j^{ols}\right|^{-\gamma}$ with $\gamma > 1$, $j = 1, 2, \ldots, p$;

4: Call

a: OAL algorithm to solve

$$\tilde{\alpha} = \arg\min_{\alpha} \left[\sum_{i=1}^{n} \left\{ -a_i (x_i^T \alpha) + \log\left(1 + e^{x_i^T \alpha}\right) \right\} + \lambda \sum_{j=1}^{p} \tilde{w}_j |\alpha_j| \right];$$

b: GOAL algorithm to solve

$$\tilde{\alpha} = \arg\min_{\alpha} \left[\sum_{i=1}^{n+p} \left\{ -a_i^* (x_i^{*T} \alpha) + \log\left(1 + e^{x_i^{*T} \alpha}\right) \right\} + \lambda_1 \sum_{j=1}^{p} \tilde{w}_j |\alpha_j| \right];$$

6: Compute predicted probabilities of exposure: $\tilde{\pi}_i \{X_i, \tilde{\alpha}\}$;

7: Compute IPT weights: $\tilde{\tau}_i = \frac{A_i}{\tilde{\pi}_i\{X_i,\tilde{\alpha}\}} + \frac{1-A_i}{1-\tilde{\pi}_i\{X_i,\tilde{\alpha}\}}$;

8: Compute direct effect: $\widehat{ACDE} = \frac{\sum_{i=1}^{n} \tilde{\tau}_i Y_i^{res} A_i}{\sum_{i=1}^{n} \tilde{\tau}_i A_i} - \frac{\sum_{i=1}^{n} \tilde{\tau}_i Y_i^{res} (1-A_i)}{\sum_{i=1}^{n} \tilde{\tau}_i (1-A_i)}$;

9: Compute indirect effect: $\widehat{AIE} = \widehat{ATE} - \widehat{ACDE}$.

3.2.4 MOAR extension for intermediate confounders (MOARI)

The diagram presented in Figure 3.1 does not consider the possibility that confounders of the mediator-outcome relationship be affected by the treatment A, that is, it hypothesizes the absence of intermediate confounding (Acharya *et al.*, 2016; Imai *et al.*, 2010c). However, this assumption may be unrealistic in several mediation contexts in practice (Acharya *et al.*, 2016); see Figure 3.2 for a visual representation of intermediate confounding. Assuming that the intermediate confounders (Z) are measured, it is possible to estimate the ACDE from our MOAR approach by making a small modification to the outcome model fitted in step 1 of the second stage of the algorithm (see Algorithm

MOARI). A similar strategy was previously adopted by Vansteelandt (2009).

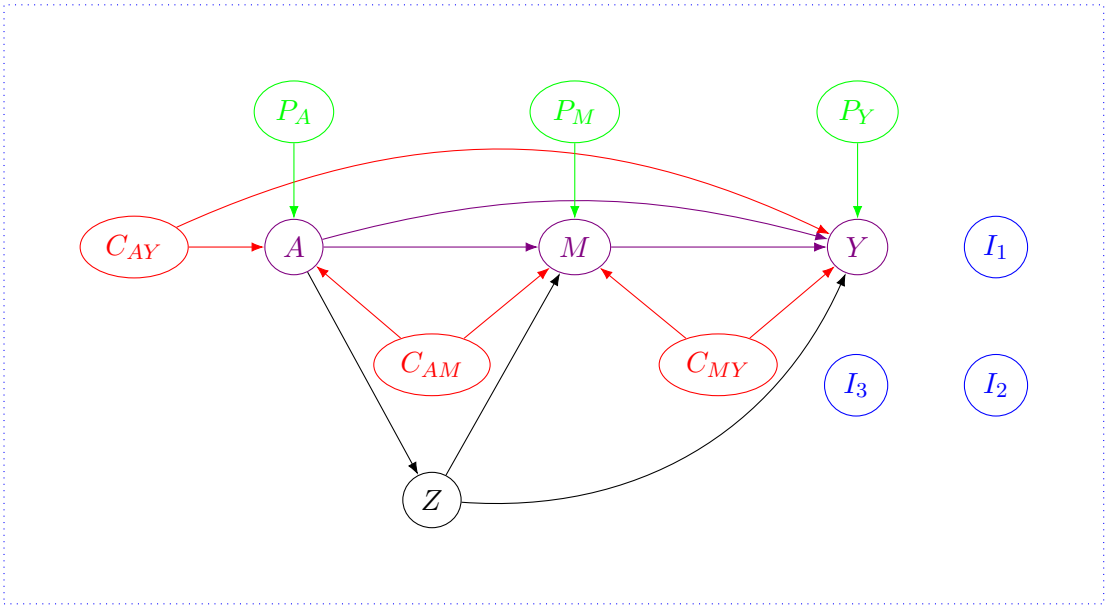


Figure 3.2 Hypothesized causal mediation graph with intermediate confounders Z.

Algorithm MOARI Part II: Estimation of average direct and indirect effects (ACDE and AIE) with intermediate confounding

1: Compute $(\tilde{\beta}_A^{ols}, \tilde{\beta}_Z^{ols}, \tilde{\beta}_M^{ols}, \tilde{\beta}^{ols}) = \arg\min_{(\beta_A,\beta_M,\beta)} \|Y - \beta_A A - \beta_Z Z - \beta_M M - \mathbf{X}\beta\|_2^2$;

2: Define $Y^{res} = Y - \tilde{\beta}_M^{ols} M$;

3: Set $\tilde{w}_j = \left|\tilde{\beta}_j^{ols}\right|^{-\gamma}$ with $\gamma > 1$, $j = 1, 2, \ldots, p$;

4: Call

a: OAL algorithm to solve

$$\tilde{\alpha} = \arg\min_{\alpha} \left[\sum_{i=1}^{n} \left\{ -a_i (x_i^T \alpha) + \log\left(1 + e^{x_i^T \alpha}\right) \right\} + \lambda \sum_{j=1}^{p} \tilde{w}_j |\alpha_j| \right];$$

b: GOAL algorithm to solve

$$\tilde{\alpha} = \arg\min_{\alpha} \left[\sum_{i=1}^{n+p} \left\{ -a_i^* (x_i^{*T} \alpha) + \log\left(1 + e^{x_i^{*T} \alpha}\right) \right\} + \lambda_1 \sum_{j=1}^{p} \tilde{w}_j |\alpha_j| \right];$$

6: Compute predicted probabilities of exposure: $\tilde{\pi}_i \{X_i, \tilde{\alpha}\}$;

7: Compute IPT weights: $\tilde{\tau}_i = \frac{A_i}{\tilde{\pi}_i\{X_i,\tilde{\alpha}\}} + \frac{1-A_i}{1-\tilde{\pi}_i\{X_i,\tilde{\alpha}\}}$;

8: Compute direct effect: $\widehat{ACDE} = \frac{\sum_{i=1}^{n} \tilde{\tau}_i Y_i^{res} A_i}{\sum_{i=1}^{n} \tilde{\tau}_i A_i} - \frac{\sum_{i=1}^{n} \tilde{\tau}_i Y_i^{res} (1-A_i)}{\sum_{i=1}^{n} \tilde{\tau}_i (1-A_i)}$;

9: Compute indirect effect: Compute indirect effect: $\widehat{AIE} = \widehat{ATE} - \widehat{ACDE}$.

3.3 Simulation study

In this section, we present the simulation study conceived for examining the performance of the proposed MOAR approach (MOAL and MGOAL) for estimating the ACDE and AIE. The main simulation scenarios described next suppose no intermediate confounding; additional scenarios presented subsequently consider a single intermediate confounder.

3.3.1 Simulation design

We describe the simulation setup used for generating the data $(\mathbf{X}, A, M, Y)$, denoting the design matrix, treatment, mediator and outcome, respectively. All observations (X_i, A_i, M_i, Y_i), $i = 1, \ldots, n$, were generated independently. We generated the vector $X_i = (X_{i1}, X_{i2}, \ldots, X_{ip})$ from a multivariate standard Gaussian distribution with pairwise correlation $\rho = 0$. The binary treatment A_i was generated from a Bernoulli distribution with $logit\{P(A_i = 1|X)\} = \sum_{j=1}^{p} \alpha_j X_{ij}$. The continuous mediator M_i was generated as $M_i = \theta_A A_i + \sum_{j=1}^{p} \theta_j X_{ij} + \epsilon_i^M$, $\epsilon_i^M \sim N(0, 1)$ and the continuous outcome Y_i as $Y_i = \beta_A A_i + \beta_M M_i + \sum_{j=1}^{p} \beta_j X_{ij} + \epsilon_i^Y$, $\epsilon_i^Y \sim N(0, 1)$. The true ACDE value was $\beta_A = 1.5$, the true AIE was $\beta_M \times \theta_A = 2 \times 1 = 2$ and the true ATE was ACDE + AIE $= 1.5 + 2 = 3.5$.

We considered three simulation scenarios inspired by those proposed by Ye et al. (2021). Our three scenarios considered common values for the mediator regression coefficients $\theta \in \mathbb{R}^p$ and treatment regression coefficients $\alpha \in \mathbb{R}^p$, but differed by the magnitude of the outcome regression coefficients $\beta \in \mathbb{R}^p$. The simulation scenarios are defined as follows:

1. Weakly related (β) : $\beta = (0.2, 0.2, 0, 0.2, 0, 0, 0, \ldots, 0)$, $\theta = (0, 1, 1, 0, 0, 1, 0, \ldots, 0)$, $\alpha = (1, 0, 0.4, 0, 1, 0, 0, \ldots, 0)$;

2. Moderately related (β): $\beta = (0.6, 0.6, 0, 0.6, 0, 0, 0, \ldots, 0)$, $\theta = (0, 1, 1, 0, 0, 1, 0, \ldots, 0)$, $\alpha = (1, 0, 0.4, 0, 1, 0, 0, \ldots, 0)$;

3. Strongly related (β) : $\beta = (1, 1, 0, 1, 0, 0, 0, \ldots, 0)$, $\theta = (0, 1, 1, 0, 0, 1, 0, \ldots, 0)$, $\alpha = (1, 0, 0.4, 0, 1, 0, 0, \ldots, 0)$.

In each scenario, the first covariate is a treatment-outcome confounder, the second is a mediator-outcome confounder, the third is a treatment-mediator confounder, the fourth covariate is a pure predictor of outcome, the fifth covariate is a pure predictor of treatment, the sixth covariate is a pure predictor of mediator and the rest of covariates are spurious ($p = 20$; $p - 6 = 14$). We simulated 1000 data sets of sizes $n = 200, 500, 1000$ for each scenario.

We compared nine IPTW estimators which differed according to how covariates were included in the PS models for the ATE and ACDE. The first four estimators require the knowledge of the covariates types (that is, confounder, pure predictor or spurious), while the last four construct the PS models upon the application of OAL, MOAL, GOAL and MGOAL, respectively. We also considered an IPTW estimator that includes all available covariates in both PS models. These nine estimators were applied to every data set of each simulation scenario, and all of them implement steps 1-2 of the Algorithm's Part II for the estimation of the ACDE. We describe the nine PS modeling strategies:

1. Targ: includes only desirable covariates (confounders and predictors) required for the unbiased and efficient estimation of the ATE and ACDE, respectively. More precisely, this strategy includes covariates $C_{AY}, C_{MY}, C_{AM}, P_Y, P_M$ in the PS model for the ATE while covariates C_{AY}, C_{MY}, P_Y are included in the PS model for the ACDE.

2. Targn: includes the confounders and the predictors of the mediator and outcome in a common adjustment set. More precisely, this strategy includes covariates $C_{AY}, C_{MY}, C_{AM}, P_Y, P_M$ in both PS models.

3. Conf: includes all confounders required for the unbiased estimation of the ATE and ACDE, respectively. More precisely, this strategy includes covariates C_{AY} and C_{AM} in the PS model for the ATE while covariates C_{AY} and C_{MY} are included in the PS model for the ACDE.

4. PotConf: includes the confounders and predictors in a common adjustment set. More precisely, this strategy includes covariates $C_{AY}, C_{MY}, C_{AM}, P_Y, P_M, P_A$ in both PS models.

5. All: includes all p available covariates in both PS models.

6. OAL: includes the covariates selected by the standard OAL algorithm (Shortreed et Ertefaie, 2017) (that is, covariates selected in step 4a of Part I of the Algorithm) in both PS models.

7. MOAL: includes the covariates selected by the proposed MOAR approach using OAL in step 4a of Parts I and II of the Algorithm.

8. GOAL: includes the covariates selected by the standard (naive) GOAL algorithm (Baldé *et al.*, 2022) (that is, covariates selected in step 4b of Part I of the Algorithm) in both PS models.

9. MGOAL: includes the covariates selected by the proposed MOAR approach using (naive) GOAL in step 4b of Parts I and II of the Algorithm.

We clarify that the difference between OAL and MOAL (or between GOAL and MGOAL) is that for OAL (GOAL) the same set of covariates, that is the one selected for the ATE, is used for estimating the ACDE, while generally different sets of covariates are used by MOAL (MGOAL).

We compared all IPTW estimators for the ATE, ACDE and AIE based on bias, standard error (SE), root mean squared error (RMSE), coverage probability and confidence interval length. For variable selection, we used the percentage of times each covariate was selected for inclusion in the PS models (tolerance = 10^{-8}; the variable is selected when its coefficient is greater than 10^{-8}). We used 200 bootstrap resamples to construct 95% normal confidence intervals.

3.3.2 Additional simulations with intermediate confounding

A similar simulation setup was used for generating the data $(\mathbf{X}, A, Z, M, Y)$, denoting the design matrix, treatment, intermediate confounder, mediator and outcome, respectively. Again, we simulated $X_i = (X_{i1}, X_{i2}, \ldots, X_{ip})$ from a multivariate standard Gaussian distribution with pairwise correlation $\rho = 0$ and binary treatment A_i from a Bernoulli distribution with $logit\{P(A_i = 1)\} = \sum_{j=1}^{p} \alpha_j X_{ij}$. A continuous intermediate confounder Z_i was generated as $Z_i = \delta_A A_i + \epsilon_i^Z$, $\epsilon_i^Z \sim N(0,1)$. The continuous mediator M_i was generated as $M_i = \theta_A A_i + \theta_Z Z_i + \sum_{j=1}^{p} \theta_j X_{ij} + \epsilon_i^M$, $\epsilon_i^M \sim N(0,1)$ and the continuous outcome Y_i as $Y_i = \beta_A A_i + \beta_Z Z_i + \beta_M M_i + \sum_{j=1}^{p} \beta_j X_{ij} + \epsilon_i^Y$, $\epsilon_i^Y \sim N(0,1)$. In addition, we set $\delta_A = \theta_Z = \beta_Z = 0.6$. The true ACDE was $\beta_A + \delta_A \times \beta_Z = 1.5 + 0.6 \times 0.6 = 1.86$, the true AIE was $\beta_M \times \theta_A + \delta_A \times \theta_Z \times \beta_M = 2 \times 1 + 0.6 \times 0.6 \times 2 = 2.72$, and the

true ATE was ATE = ACDE + AIE = $1.86 + 2.72 = 4.58$. We also considered the same three sets of values as in the main simulation scenarios for β, θ and α. The same nine IPTW estimators were also evaluated for the ATE, ACDE and AIE.

3.3.3 Results

Tables 3.1-3.3 and Figure 3.3 present results associated with Scenarios 1-3 of the main simulations, displaying the bias, SE, RMSE, coverage probability and confidence interval length of the nine estimators IPTW for the ATE, ACDE and AIE, namely Targ, Targn, Conf, PotConf, All, OAL, MOAL, GOAL, and MGOAL.

For Scenario 1 when $n = 200$, all estimators showed relatively small biases for the ATE, with biases slightly larger for Conf and the data-driven selection approaches (OAL, MOAL, GOAL and MGOAL) (see Table 3.1). When $n = 500$ and $n = 1000$, Targ, Targn, Conf and PotConf, and All exhibited very small biases in the ATE. The biases of OAL, MOAL, GOAL and MGOAL estimators for the ATE decreased as n increased to 1000. At all sample sizes considered, the data-driven approaches showed a variance similar to the Targ and Targn estimators for the ATE. The All estimator exhibited the largest variability for the ATE, while the estimators Conf and PotConf were between Targ(n) and All. The coverage probabilities for the ATE were close to 0.95 throughout, except for All at the smaller sample size ($n = 200$), for which the coverage probability was estimated at 1. For the ACDE, all estimators had very small or no bias, except OAL and GOAL which showed larger biases especially at the smallest sample size. It should be noted that confounders C_{MY} are judged desirable covariates by OAL and GOAL solely based on their role as predictors of the outcome, and not because of their role as confounders. However recall that C_{MY} are required for unbiased estimation of the ACDE. The merit of MOAL and MGOAL, as compared to OAL and GOAL, for the ACDE is clear from the results since the formers are systematically found less biased and variable than latters, with bias and variance of similar magnitude than found for Targ(n). No notable differences in performance were observed when comparing MOAL and MGOAL for the ACDE. For the AIE, an opposite pattern of results was found for the data-driven approaches, as OAL and GOAL were observed less biased and variable

than MOAL and MGOAL, especially at the smallest sample size ($n = 200$).

We observed similar results in Scenarios 2 and 3 (see Tables 3.2-3.3), with the estimators exhibiting overall slightly more bias and variability. For the smallest sample size ($n = 200$), GOAL/MGOAL biases for the ATE and ACDE were found noticeably smaller as compared to OAL/MOAL.

Figure 3.3 presents the covariates' probabilities of inclusion as obtained by the proposed MOAR approach (MOAL and MGOAL) as well as for OAL and GOAL, with a focus on the ACDE. For this effect, we found that MOAR included treatment-outcome (C_{AY}) and mediator-outcome (C_{MY}) confounders and the pure predictor of the outcome (P_Y) in the PS model for the ACDE with a large probability, while a much smaller probability of inclusion was obtained for the other covariates. As n increased, covariates C_{AY}, C_{MY} and P_Y were selected in an increasingly larger proportion of times, while the proportion of times that other covariates (C_{AM}, P_M, P_A and spurious) were selected declined. As can be seen through the OAL/GOAL results in Figure 3.3, we found, as expected, that the treatment-mediator confounder C_{AM} and the pure predictor of the mediator P_M were selected with high probability by all four data-driven approaches for the ATE. Overall, MOAR performed well in terms of variable selection by selecting all true desirable covariates in both PS models. The two versions of MOAR, namely MOAL and MGOAL, had similar performance in terms of covariate selection rate. In Web Appendix B (see Section 3.6.2), we illustrate the impact of computing the adaptive weights from $(\tilde{\beta}_A^{ols}, \tilde{\beta}_M^{ols}, \tilde{\beta}^{ols}) = \arg\min_{(\beta_A,\beta_M,\beta)} \|Y - \beta_A A - \beta_M M - \mathbf{X}\beta\|_2^2$ instead of $(\hat{\beta}_A^{ols}, \hat{\beta}^{ols}) = \arg\min_{(\beta_A,\beta)} \|Y - \beta_A A - \mathbf{X}\beta\|_2^2$ when targeting the ATE (refer to step 2 in Algorithm MOAR Part I). As alluded to in the introduction, building the adaptive weights based on an outcome model that includes the mediator prevents OAL to select confounders C_{AM} since these predict the treatment but not the outcome after adjustment on M (see Web Figure 3.5). This is undesirable since confounders C_{AM} are required for unbiased effect estimation (see Web Appendix B in Section 3.6.2 and Web Tables 3.7-3.9). Note that unlike confounders C_{AM}, it can be expected that common causes of the treatment, mediator and outcome (that is, confounders C_{AMY}) would not be excluded by OAL since C_{AMY} would still predict the outcome after adjustment for M.

Table 3.4 present results associated with Scenario 1 in the presence of an intermediate confounder Z. Due to space constraints, the results associated with Scenarios 2-3 are presented in Web Appendix C in Section 3.6.3 (see Web Tables 3.10-3.11 and Web Figure 3.6). Simulation results showed that MOARI had similar estimation and covariate selection patterns than MOAR.

3.4 Real data application

Marijuana use is common among high school and college students (Meier *et al.*, 2015; Arria *et al.*, 2015; NIH, 2020). Previous studies have established an association between marijuana use and poorer academic performance among high school and college students (Meier *et al.*, 2015; Arria *et al.*, 2015). One hypothesis that has been put forward is that missing classes may explain the association between marijuana use and lower grades (Meier *et al.*, 2015; Arria *et al.*, 2015). We used data collected from the Harvard School of Public Health College Alcohol Study (round 1999) to illustrate our MOAR approach while assessing the potential mediating role of school disengagement in the association between marijuana use and grade point average (GPA). This study collected information on students and their college environment: demographics, substance use, academic behaviors, grades, personal problems and campus policies. Our example is based on a sample of $n = 918$ undergraduate and graduate students with complete measurements on $p = 10$ selected covariates, in addition to exposure, mediator and outcome variables. We took *marijuana use* (yes/no) in the past 12 months as the exposure variable, *school disengagement* since beginning of the school year as the mediator variable and current year *GPA* as the outcome variable. The mediator was defined as the sum of the two ordered categorical variables *missed class* and *behind in schoolwork*. We refer the reader to Web Appendix D in Section 3.6.4 for more details on covariates.

A fully adjusted mediation analysis was initially done using the product method. Note that results obtained from the product method are equivalent to those from the difference method under standard linear models for the outcome and mediator. The corresponding estimated path coefficients are presented in Figure 3.4.

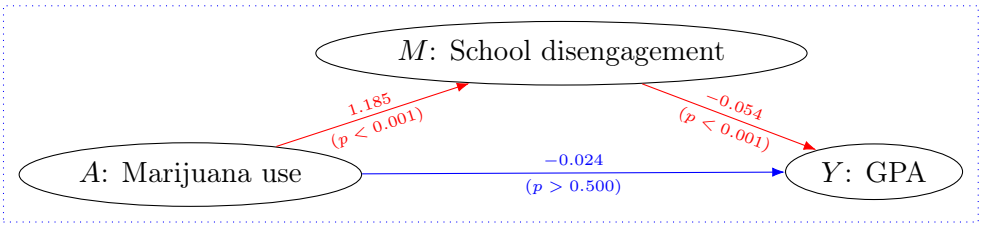


Figure 3.4 Simple mediation graph displaying the effect of marijuana use in past 12 months (exposure) through school disengagement (mediator) since the beginning of the school year on the current year grade point average (GPA; outcome) among undergraduate and graduate students with known demographic risk factors and baseline information on substance use and college academic behaviors ($n = 918$). P-values were obtained from a fully adjusted mediation analysis with the product method.

We obtained a statistically significant indirect effect of marijuana use (exposure) on GPA (outcome) via school disengagement (mediator) of -0.064 with a 95% CI of $(-0.095, -0.040)$, while the direct effect of marijuana use on GPA was not found statistically significant (-0.024, 95% CI: $(-0.115, 0.060)$). However, we obtained a statistically significant total effect of marijuana use on GPA (-0.089, 95% CI: $(-0.177, -0.010)$).

Tables 3.5 and 3.6 present the results when data-driven approaches OAL, GOAL and MOAR (MOAL and MGOAL) as well as the All approach were used to fit the PS models to estimate the ATE and ACDE. We used bootstrap with 10 000 resamples to construct both 95% normal confidence intervals (95% NCI) and 95% percentile confidence intervals (95% PCI) for the effects. For each bootstrap resample, we verified which covariates were selected for inclusion in the PS models (tolerance $= 10^{-8}$).

Results are presented in Tables 3.5 and 3.6. With all covariates included in both PS model (All approach), we found an IPTW estimate for the ATE of -0.115 with a 95% PCI of $(-0.205, -0.011)$. The estimated ATE using OAL/MOAL was essentially the same as for GOAL/MGOAL, that is (OAL/MOAL): -0.078, 95% PCI: $(-0.188, 0.012)$; GOAL/MGOAL: -0.079, 95% PCI: $(-0.192, 0.010)$. With all covariates, the estimated ACDE was -0.050 with a 95% PCI of $(-0.160, 0.048)$. The estimated ACDE (point and interval estimates) for OAL and GOAL were found very similar (OAL: -0.004, 95%

PCI: $(-0.128, 0.075)$; GOAL: -0.004, 95% PCI: $(-0.137, 0.072)$. The estimated ACDE for MOAL and MGOAL were also very similar (MOAL: -0.028, 95% PCI: (-0.123, 0.063); MGOAL: -0.031, 95% PCI: $(-0.140, 0.059)$). The ACDE estimates obtained from MOAL and MGOAL were in greater agreement with the product method's results than those obtained with OAL/GOAL. MOAL and MGOAL also showed the shortest confidence intervals for the ACDE.

The estimated AIE (point and interval estimates) for OAL and GOAL were found very similar (OAL: -0.075, 95% PCI: $(-0.099, -0.042)$; GOAL: -0.074, 95% PCI: $(-0.098, -0.041)$). The estimated AIE for MOAL and MGOAL were also very similar (MOAL: -0.050, 95% PCI: $(-0.099, -0.029)$; MGOAL: -0.048, 95% PCI: $(-0.099, -0.018)$). Contrary to the ACDE, the AIE estimates obtained from OAL and GOAL were in greater agreement with the product method's results than those obtained with MOAL/MGOAL. OAL and GOAL also showed the shortest confidence intervals for the AIE.

In Table 3.6, we present MOAL and MGOAL's probabilities of inclusion for each covariate in the PS models associated with the ATE and ACDE. Note that the estimated probabilities shown for the ATE also correspond to those obtained for OAL and GOAL since OAL and MOAL (resp. GOAL and MGOAL) are equivalent for estimating the ATE. For a given effect, MOAL and MGOAL selected covariates with similar probabilities. For both MOAL and MGOAL, a difference in probability of inclusion greater than 10% between the ATE and ACDE was observed for variables *Smoke cigarettes*, *Education satisfaction* and *Alcohol drinking and binge*. Otherwise both approaches selected the remaining variables at similar rates for the ATE and ACDE. Variable selection results showed that variable *Gender* was almost always selected for both ATE and ACDE, while covariate *High school athletics* was selected only $\approx$18% of times for ATE and ACDE.

3.5 Discussion

We introduced a new regularized variable selection approach for causal mediation analysis. Both outcome adaptive lasso (OAL) (Shortreed et Ertefaie, 2017) and generalized outcome adaptive lasso (GOAL) (Baldé *et al.*, 2022) have been previously found successful for variable selection by targeting covariates that yield an unbiased and efficient IPTW estimator for the ATE. However, data-driven causal variable selection approaches such as OAL and GOAL are not intrinsically designed for selecting appropriate covariates for estimating direct and indirect effects. In particular, OAL and GOAL may miss mediator-outcome confounders and include pure predictors of the mediator although those are respectively necessary and undesirable when targeting the ACDE. Instead of selecting a single set of covariates as for OAL and GOAL, our proposed variable selection approach for mediation MOAR is designed to select two separate sets of covariates, each tailored for its targeted effect.

Our simulation results suggest that both versions of MOAR (MOAL and MGOAL) improve on OAL and GOAL when estimating the ACDE, especially when the sample size is small. Moreover, our results showed that the performance of all data-driven estimators (OAL, MOAL, GOAL, MGOAL) was relatively close to the one of estimator Targ in terms of estimation accuracy in all scenarios. Recall that Targ requires perfect knowledge of covariate types: for the ATE it includes all treatment-outcome confounders (including treatment-mediator confounders) and all outcome predictors (including predictors of the mediator) for the ATE, while for the ACDE it includes mediator-outcome confounders, treatment-outcome confounders (that are not treatment-mediator confounders) and predictors of the outcome (that are not predictors of the mediator). In our simulations, we found that estimator MOAL performed globally as MGOAL for all effects (ATE, ACDE and AIE), but the latter approach was found noticeably less biased, in particular with smaller sample sizes. In terms of variable selection, MOAL and MGOAL performed equivalently at all sample sizes in all scenarios investigated. The extension of MOAR for dealing with intermediate confounding (MOARI) was found working satisfactorily in the simulation scenarios investigated.

A current limitation of MOAR is that the outcome model excludes an interaction term between treatment and mediator. A potential extension of our proposed approach would therefore be to allow for such a term in the outcome model. In this work, we evaluated MOAR's performance in simple scenarios with a relatively small number of independant covariates. However we expect, owing to OAL and GOAL properties, that MOAR can be used in high-dimensional settings and/or with correlated covariates. With pronounced collinearity in covariates, we recommend implementing MOAR with the iterative version of GOAL (Baldé *et al.*, 2022), as opposed to the naive GOAL or OAL. We leave the high-dimensional scenarios and oracle properties of MOAR for future investigations.

3.6 Web Appendix

3.6.1 Web Appendix A. Sequential g-estimation [Stijn Vansteelandt, 2009]

In this section, we briefly recall the sequential g-estimation method. Sequential g-estimation is a simple and efficient method that allows direct modeling of the average controlled direct effect, and which is designed to accommodate intermediate confounding. In the sequel, we present the three steps involved in sequential g-estimation:

1. First, estimate the effect of M on Y using the following regression:

$$(\hat{\beta}_A^{ols}, \hat{\beta}_M^{ols}, \hat{\beta}^{ols}) = \arg \min_{(\beta_A, \beta_M, \beta)} \|Y - \beta_A A - \beta_M M - \mathbf{X}\beta\|_2^2$$

   or (in the presence of intermediate confounders)

$$(\hat{\beta}_A^{ols}, \hat{\beta}_Z^{ols}, \hat{\beta}_M^{ols}, \hat{\beta}^{ols}) = \arg \min_{(\beta_A, \beta_M, \beta)} \|Y - \beta_A A - \beta_Z Z - \beta_M M - \mathbf{X}\beta\|_2^2 ;$$

2. Second, form a new outcome variable Y_{res} free of the effect of M:

$$Y_{res} = Y - \hat{\beta}_M^{ols} M;$$

3. Third, the average controlled direct effect is encoded in the γ_A coefficient in :

$$Y_{res} = \gamma_A A + \gamma_X X + \nu, \quad \nu \sim N(0, \sigma_\nu^2 \mathbf{I}),$$

   and estimated accordingly.

3.6.2 Web Appendix B. Impact of adaptive weights specification

Web Figure 3.5 and Tables 3.7-3.9 present OAL, MOAL, GOAL and MGOAL results associated to Scenarios 1-3 with adaptive weights computed using the covariates' unpenalized coefficient estimates obtained by fitting a linear regression of Y on A, M and X. This way to compute weights mimic that proposed by Ye et al. (2021). As seen in Web Figure 3.5, covariates C_{AM} and P_M are underselected as compared to when adaptive weights computed using the covariates' unpenalized coefficient estimates obtained by fitting a linear regression of Y on A and X (compare with OAL/GOAL lines in Figure 3.3). In Web Tables 3.7-3.9, non negligible biases and increased variances for the ATE and AIE are obtained for all data-driven approaches with this alternative strategy for defining weights.

3.6.3 Web Appendix C. Additional simulation results with intermediate confounding

Web Tables 3.10-3.11 and Web Figure 3.6 present results associated to Scenarios 2-3 with intermediate confounding.

3.6.4 Web Appendix D. Description of variables

In this section, we define the ten variables that were used in our real application based on the Harvard School of Public Health College Alcohol Study (round 1999). More specifically we detail the variables according the study's Codebook:

1. Gender (male: 1 / female: 0) is variable A2 (gender). Gender takes the value 1 for males and 0 for females.

2. Race (White: 1 / otherwise: 0) is obtained from the variable race with four categories. We defined Race in two categories: Race takes the value 1 if the student is White and 0 otherwise (all others races).

3. Smoke cigarettes (yes: 1 / no: 0) is obtained from the variable E1N. We redefined E1N in two categories: Smoke cigarettes takes the value 1 if the student smokes cigarettes and 0 if not.

4. Education satisfaction (yes: 1 / otherwise: 0) is obtained from the variable F1 with four categories. We redefined F1 in two categories: Education satisfaction takes the value 1 if the student is very or somewhat satisfied with education and 0 otherwise.

5. Parents education with a 4 years college diploma or more (yes: 1 / no: 0) is obtained from the variables G18 (father schooling) and G19 (mother schooling). Parents education takes the value 1 if the father or the mother has a 4 years college diploma or more and 0 if not.

6. Alcohol drinking and binge (yes: 1 / no: 0) is the variable hsubinge. Alcohol drinking and binge takes the value 1 if the student drank and binged during his / her last year in high school and 0 if not.

7. Member of fraternity / sorority (yes : 1 / no: 0) is the covariate memgreek. Member of fraternity / sorority takes the value 1 if the student is a member of a fraternity or a sorority and 0 if not.

8. Parents approval of drinking (yes : 1 / otherwise : 0) is obtained from the variable G17 with four categories. We redefined G17 in two categories. Parents approval of drinking takes the value 0 if the family of the student did not approve of drinking and 1 otherwise.

9. Number of other drug use is obtained from the variables E1B-E1M and E1O-E1P. The variable Number of other drug use counts the number of other drug use by

combining all the variables E1B-E1M and E1O-E1P. Each of the covariates E1B-E1M and E1O-E1P was dichotomized (1 / 0) before the combination.

10. High school athletics (yes: 1 / no: 0) is obtained from the variable G14 with three categories. We redefined High school athletics in two categories: High school athletics takes the value 1 if the student participated in high school athletics (with or without varsity letter) and 0 if not.

Tableau 3.1 Bias, standard error (SE), root mean squared error (RMSE) and coverage probability (confidence interval length) of the nine studied inverse probability of treatment weighting estimators (Targ, Targn, Conf, PotConf, ALL, OAL, MOAL, GOAL, MGOAL) for the average total effect (ATE), average controlled direct effect (ACDE) and average indirect effect (AIE) under Scenario 1, with sample sizes $n = 200, 500, 1000$ (results based on 1000 estimates).

	ATE				**ACDE**				**AIE**			
	Bias	**SE**	**RMSE**	**P (Len)**	**Bias**	**SE**	**RMSE**	**P (Len)**	**Bias**	**SE**	**RMSE**	**P (Len)**
$n = 200$												
Targ	0.03	0.40	0.41	0.95 (1.67)	0.00	0.19	0.19	0.96 (0.75)	0.02	0.39	0.39	0.95 (1.60)
Targn	0.03	0.40	0.41	0.95 (1.67)	0.00	0.18	0.18	0.96 (0.72)	0.02	0.37	0.37	0.95 (1.55)
Conf	0.06	0.63	0.64	0.93 (2.34)	0.00	0.19	0.19	0.96 (0.75)	0.05	0.62	0.63	0.93 (2.31)
PotConf	0.04	0.62	0.62	0.96 (2.32)	0.01	0.22	0.22	0.94 (0.84)	0.03	0.56	0.56	0.96 (2.17)
All	0.04	0.79	0.79	1.00 (4.13)	0.01	0.25	0.25	0.98 (1.18)	0.03	0.74	0.74	1.00 (3.89)
OAL	0.09	0.40	0.41	0.95 (1.61)	0.08	0.19	0.21	0.93 (0.76)	0.01	0.35	0.35	0.96 (1.46)
MOAL	0.09	0.40	0.41	0.95 (1.61)	0.02	0.19	0.19	0.96 (0.79)	0.06	0.38	0.39	0.96 (1.58)
GOAL	0.06	0.41	0.42	0.95 (1.64)	0.08	0.20	0.21	0.93 (0.77)	-0.01	0.36	0.36	0.96 (1.48)
MGOAL	0.06	0.41	0.42	0.95 (1.64)	0.01	0.19	0.19	0.96 (0.81)	0.05	0.39	0.39	0.96 (1.60)
$n = 500$												
Targ	0.01	0.24	0.24	0.95 (0.98)	-0.00	0.11	0.11	0.95 (0.46)	0.02	0.23	0.23	0.96 (0.93)
Targn	0.01	0.24	0.24	0.95 (0.98)	-0.00	0.11	0.11	0.96 (0.44)	0.01	0.22	0.22	0.96 (0.91)
Conf	0.02	0.37	0.37	0.94 (1.48)	0.00	0.12	0.12	0.95 (0.46)	0.02	0.37	0.37	0.94 (1.45)
PotConf	0.01	0.39	0.39	0.94 (1.40)	-0.00	0.13	0.13	0.95 (0.52)	0.02	0.36	0.36	0.94 (1.30)
All	0.01	0.43	0.43	0.97 (1.74)	-0.00	0.14	0.14	0.96 (0.57)	0.01	0.40	0.40	0.97 (1.62)
OAL	0.05	0.25	0.26	0.96 (1.02)	0.05	0.13	0.14	0.93 (0.51)	-0.00	0.22	0.22	0.96 (0.91)
MOAL	0.05	0.25	0.26	0.96 (1.02)	0.00	0.11	0.11	0.96 (0.48)	0.05	0.24	0.25	0.95 (0.98)
GOAL	0.04	0.25	0.25	0.96 (1.03)	0.05	0.13	0.14	0.94 (0.52)	-0.01	0.22	0.22	0.96 (0.91)
MGOAL	0.04	0.25	0.25	0.96 (1.03)	-0.00	0.12	0.12	0.96 (0.49)	0.04	0.24	0.25	0.95 (0.99)
$n = 1000$												
Targ	0.01	0.18	0.18	0.94 (0.68)	0.00	0.08	0.08	0.94 (0.32)	0.01	0.17	0.17	0.94 (0.65)
Targn	0.01	0.18	0.18	0.94 (0.68)	0.00	0.08	0.08	0.94 (0.31)	0.01	0.16	0.16	0.94 (0.63)
Conf	0.02	0.27	0.27	0.94 (1.04)	0.00	0.09	0.09	0.95 (0.33)	0.01	0.26	0.26	0.94 (1.02)
PotConf	0.00	0.28	0.28	0.94 (1.01)	-0.00	0.10	0.10	0.93 (0.37)	0.00	0.26	0.26	0.95 (0.94)
All	0.00	0.29	0.29	0.95 (1.11)	-0.00	0.11	0.11	0.93 (0.38)	0.00	0.27	0.27	0.96 (1.03)
OAL	0.03	0.18	0.19	0.95 (0.74)	0.03	0.10	0.10	0.93 (0.39)	0.00	0.16	0.16	0.97 (0.65)
MOAL	0.03	0.18	0.19	0.95 (0.74)	0.00	0.08	0.08	0.95 (0.33)	0.03	0.18	0.18	0.96 (0.71)
GOAL	0.02	0.18	0.18	0.96 (0.74)	0.03	0.10	0.10	0.94 (0.40)	-0.01	0.16	0.16	0.97 (0.65)
MGOAL	0.02	0.18	0.18	0.96 (0.74)	-0.00	0.08	0.08	0.96 (0.34)	0.02	0.18	0.18	0.96 (0.71)

Tableau 3.2 Bias, standard error (SE), root mean squared error (RMSE) and coverage probability (confidence interval length) of the nine studied inverse probability of treatment weighting estimators (Targ, Targn, Conf, PotConf, ALL, OAL, MOAL, GOAL, MGOAL) for the average total effect (ATE), average controlled direct effect (ACDE) and average indirect effect (AIE) under Scenario 2, with sample sizes $n = 200, 500, 1000$ (results based on 1000 estimates).

	ATE				**ACDE**				**AIE**			
	Bias	**SE**	**RMSE**	**P (Len)**	**Bias**	**SE**	**RMSE**	**P (Len)**	**Bias**	**SE**	**RMSE**	**P (Len)**
$n = 200$												
Targ	0.03	0.42	0.42	0.96 (1.73)	0.01	0.19	0.19	0.96 (0.78)	0.02	0.39	0.39	0.95 (1.64)
Targn	0.03	0.42	0.42	0.96 (1.73)	0.01	0.19	0.19	0.96 (0.77)	0.02	0.37	0.37	0.95 (1.55)
Conf	0.06	0.68	0.69	0.94 (2.53)	0.01	0.21	0.21	0.95 (0.85)	0.06	0.67	0.67	0.94 (2.47)
PotConf	0.05	0.66	0.66	0.96 (2.45)	0.01	0.26	0.26	0.94 (0.97)	0.03	0.56	0.56	0.96 (2.17)
All	0.05	0.85	0.85	1.00 (4.42)	0.02	0.31	0.31	0.99 (1.52)	0.03	0.74	0.74	1.00 (3.89)
OAL	0.10	0.43	0.44	0.94 (1.75)	0.06	0.21	0.22	0.96 (0.94)	0.04	0.37	0.38	0.96 (1.52)
MOAL	0.10	0.43	0.44	0.94 (1.75)	0.02	0.19	0.19	0.96 (0.80)	0.08	0.40	0.41	0.96 (1.70)
GOAL	0.06	0.43	0.43	0.95 (1.77)	0.05	0.21	0.22	0.97 (0.96)	0.01	0.37	0.37	0.96 (1.53)
MGOAL	0.06	0.43	0.43	0.95 (1.77)	0.00	0.19	0.19	0.97 (0.82)	0.06	0.40	0.41	0.96 (1.72)
$n = 500$												
Targ	0.02	0.24	0.24	0.95 (1.01)	-0.00	0.12	0.12	0.95 (0.47)	0.02	0.23	0.23	0.96 (0.95)
Targn	0.02	0.24	0.24	0.95 (1.01)	0.00	0.11	0.11	0.96 (0.46)	0.01	0.22	0.22	0.96 (0.91)
Conf	0.03	0.40	0.40	0.95 (1.59)	0.00	0.13	0.13	0.95 (0.52)	0.02	0.39	0.39	0.95 (1.55)
PotConf	0.02	0.41	0.41	0.94 (1.46)	-0.00	0.15	0.15	0.95 (0.59)	0.02	0.36	0.36	0.94 (1.30)
All	0.01	0.46	0.46	0.97 (1.84)	-0.00	0.16	0.16	0.97 (0.69)	0.01	0.40	0.40	0.97 (1.62)
OAL	0.04	0.26	0.26	0.96 (1.08)	0.02	0.12	0.12	0.97 (0.54)	0.02	0.23	0.23	0.96 (0.95)
MOAL	0.04	0.26	0.26	0.96 (1.08)	0.01	0.12	0.12	0.97 (0.49)	0.03	0.25	0.25	0.96 (1.03)
GOAL	0.02	0.26	0.26	0.97 (1.08)	0.01	0.12	0.12	0.97 (0.54)	0.01	0.24	0.24	0.96 (0.95)
MGOAL	0.02	0.26	0.26	0.97 (1.08)	-0.00	0.12	0.12	0.97 (0.49)	0.02	0.25	0.25	0.96 (1.03)
$n = 1000$												
Targ	0.01	0.18	0.18	0.94 (0.70)	0.00	0.09	0.09	0.95 (0.33)	0.01	0.17	0.17	0.94 (0.66)
Targn	0.01	0.18	0.18	0.94 (0.70)	0.00	0.08	0.08	0.94 (0.32)	0.01	0.16	0.16	0.94 (0.63)
Conf	0.02	0.29	0.29	0.94 (1.13)	0.00	0.10	0.10	0.94 (0.37)	0.01	0.28	0.28	0.94 (1.09)
PotConf	0.00	0.30	0.30	0.94 (1.06)	0.00	0.12	0.12	0.93 (0.42)	0.00	0.26	0.26	0.95 (0.94)
All	0.00	0.31	0.31	0.95 (1.17)	-0.00	0.12	0.12	0.93 (0.45)	0.00	0.27	0.27	0.96 (1.03)
OAL	0.02	0.18	0.18	0.95 (0.74)	0.01	0.08	0.09	0.96 (0.35)	0.01	0.16	0.17	0.96 (0.66)
MOAL	0.02	0.18	0.18	0.95 (0.74)	0.01	0.09	0.09	0.96 (0.34)	0.02	0.17	0.17	0.96 (0.71)
GOAL	0.01	0.18	0.18	0.96 (0.74)	0.00	0.09	0.09	0.96 (0.35)	0.00	0.16	0.16	0.96 (0.66)
MGOAL	0.01	0.18	0.18	0.96 (0.74)	-0.00	0.09	0.09	0.96 (0.34)	0.01	0.17	0.17	0.97 (0.71)

Tableau 3.3 Bias, standard error (SE), root mean squared error (RMSE) and coverage probability (confidence interval length) of the nine studied inverse probability of treatment weighting estimators (Targ, Targn, Conf, PotConf, ALL, OAL, MOAL, GOAL, MGOAL) for the average total effect (ATE), average controlled direct effect (ACDE) and average indirect effect (AIE) under Scenario 3, with sample sizes $n = 200, 500, 1000$ (results based on 1000 estimates).

	ATE				**ACDE**				**AIE**			
	Bias	**SE**	**RMSE**	**P (Len)**	**Bias**	**SE**	**RMSE**	**P (Len)**	**Bias**	**SE**	**RMSE**	**P (Len)**
$n = 200$												
Targ	0.03	0.43	0.43	0.96 (1.81)	0.01	0.21	0.21	0.96 (0.84)	0.02	0.40	0.40	0.95 (1.68)
Targn	0.03	0.43	0.43	0.96 (1.81)	0.01	0.21	0.21	0.96 (0.86)	0.02	0.37	0.37	0.95 (1.55)
Conf	0.07	0.74	0.74	0.94 (2.76)	0.01	0.25	0.25	0.95 (1.02)	0.06	0.71	0.71	0.94 (2.64)
PotConf	0.05	0.71	0.71	0.96 (2.61)	0.02	0.33	0.33	0.95 (1.18)	0.03	0.56	0.56	0.96 (2.17)
All	0.06	0.92	0.92	1.00 (4.78)	0.03	0.41	0.41	0.99 (2.04)	0.03	0.74	0.74	1.00 (3.89)
OAL	0.09	0.44	0.44	0.95 (1.81)	0.05	0.20	0.21	0.96 (0.95)	0.04	0.38	0.38	0.95 (1.54)
MOAL	0.09	0.44	0.44	0.95 (1.81)	0.03	0.20	0.20	0.96 (0.85)	0.06	0.41	0.41	0.96 (1.75)
GOAL	0.04	0.44	0.44	0.96 (1.84)	0.03	0.21	0.21	0.98 (0.98)	0.01	0.38	0.38	0.96 (1.57)
MGOAL	0.04	0.44	0.44	0.96 (1.84)	0.00	0.20	0.20	0.97 (0.85)	0.04	0.41	0.41	0.97 (1.79)
$n = 500$												
Targ	0.02	0.25	0.25	0.95 (1.04)	-0.00	0.12	0.12	0.95 (0.50)	0.02	0.24	0.24	0.96 (0.97)
Targn	0.02	0.25	0.25	0.95 (1.04)	0.00	0.12	0.12	0.96 (0.50)	0.01	0.22	0.22	0.96 (0.91)
Conf	0.03	0.43	0.44	0.95 (1.73)	0.00	0.16	0.16	0.95 (0.62)	0.03	0.42	0.42	0.94 (1.66)
PotConf	0.02	0.45	0.45	0.94 (1.55)	0.00	0.19	0.19	0.95 (0.71)	0.02	0.36	0.36	0.94 (1.30)
All	0.01	0.50	0.50	0.97 (1.97)	-0.00	0.21	0.21	0.98 (0.87)	0.01	0.40	0.40	0.97 (1.62)
OAL	0.04	0.26	0.27	0.97 (1.10)	0.02	0.12	0.12	0.96 (0.54)	0.02	0.23	0.24	0.96 (0.96)
MOAL	0.04	0.26	0.27	0.97 (1.10)	0.01	0.12	0.12	0.96 (0.51)	0.03	0.25	0.25	0.96 (1.05)
GOAL	0.01	0.26	0.26	0.97 (1.11)	0.01	0.12	0.12	0.97 (0.54)	0.01	0.24	0.24	0.96 (0.96)
MGOAL	0.01	0.26	0.26	0.97 (1.11)	-0.00	0.12	0.12	0.97 (0.51)	0.02	0.25	0.25	0.97 (1.06)
$n = 1000$												
Targ	0.01	0.19	0.19	0.93 (0.73)	0.00	0.09	0.09	0.95 (0.35)	0.01	0.17	0.17	0.94 (0.67)
Targn	0.01	0.19	0.19	0.93 (0.73)	0.00	0.09	0.09	0.95 (0.35)	0.01	0.16	0.16	0.94 (0.63)
Conf	0.02	0.32	0.32	0.95 (1.23)	0.01	0.11	0.11	0.94 (0.44)	0.01	0.30	0.30	0.95 (1.17)
PotConf	0.01	0.32	0.32	0.94 (1.13)	0.00	0.15	0.15	0.93 (0.50)	0.00	0.26	0.26	0.95 (0.94)
All	0.00	0.34	0.34	0.95 (1.25)	-0.00	0.16	0.16	0.94 (0.55)	0.00	0.27	0.27	0.96 (1.03)
OAL	0.02	0.19	0.19	0.95 (0.77)	0.01	0.09	0.09	0.96 (0.37)	0.01	0.16	0.16	0.96 (0.67)
MOAL	0.02	0.19	0.19	0.95 (0.77)	0.01	0.09	0.09	0.96 (0.35)	0.01	0.18	0.18	0.96 (0.72)
GOAL	0.00	0.19	0.19	0.96 (0.76)	0.00	0.09	0.09	0.97 (0.37)	0.00	0.16	0.16	0.96 (0.67)
MGOAL	0.00	0.19	0.19	0.96 (0.76)	-0.00	0.09	0.09	0.96 (0.35)	0.01	0.18	0.18	0.96 (0.72)

Figure 3.3 Probability of covariate being included in PS model for estimating the average controlled direct effect (ACDE) under Scenarios 1-3 (by row) with sample sizes $n = 200, 500, 1000$. Covariate index: 1, C_{AY}; 2, C_{MY}; 3, C_{AM}; 4, P_Y; 5, P_A; 6, P_M ; 7-20, spurious.

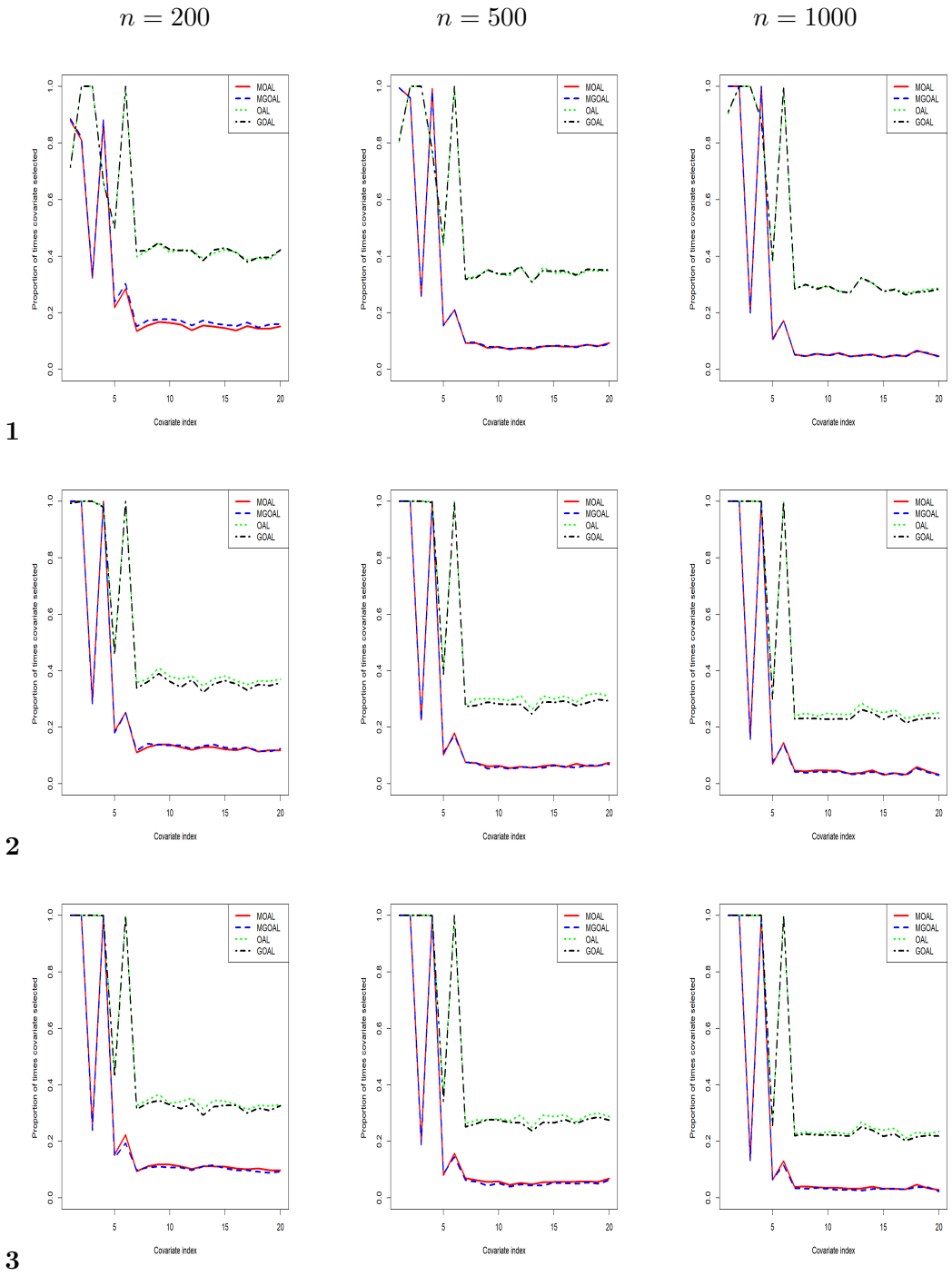


Tableau 3.4 Bias, standard error (SE), root mean squared error (RMSE) and coverage probability (confidence interval length) of the nine studied inverse probability of treatment weighting estimators (Targ, Targn, Conf, PotConf, ALL, OAL, MOAL, GOAL, MGOAL) for the average total effect (ATE), average controlled direct effect (ACDE) and average indirect effect (AIE) under Scenario 1 (in presence of intermediate confounding), with sample sizes $n = 200, 500, 1000$ (results based on 1000 estimates).

	ATE				**ACDE**				**AIE**			
	Bias	**SE**	**RMSE**	**P (Len)**	**Bias**	**SE**	**RMSE**	**P (Len)**	**Bias**	**SE**	**RMSE**	**P (Len)**
$n = 200$												
Targ	0.02	0.50	0.50	0.97 (2.04)	0.01	0.23	0.23	0.94 (0.89)	0.00	0.45	0.45	0.96 (1.84)
Targn	0.02	0.50	0.50	0.97 (2.04)	0.01	0.23	0.23	0.94 (0.86)	0.00	0.44	0.43	0.96 (1.77)
Conf	0.03	0.66	0.66	0.94 (2.61)	0.01	0.23	0.23	0.94 (0.90)	0.02	0.63	0.63	0.94 (2.46)
PotConf	0.04	0.70	0.70	0.96 (2.69)	0.02	0.28	0.28	0.94 (1.00)	0.02	0.62	0.62	0.96 (2.39)
All	0.03	0.89	0.89	1.00 (4.56)	0.02	0.33	0.33	0.98 (1.38)	0.01	0.78	0.78	1.00 (4.14)
OAL	0.07	0.49	0.49	0.95 (2.04)	0.09	0.24	0.25	0.94 (0.91)	-0.01	0.41	0.41	0.95 (1.73)
MOAL	0.07	0.49	0.49	0.95 (2.04)	0.03	0.24	0.24	0.95 (0.94)	0.04	0.45	0.45	0.95 (1.86)
GOAL	0.05	0.50	0.51	0.95 (2.07)	0.08	0.25	0.26	0.93 (0.92)	-0.03	0.42	0.42	0.96 (1.74)
MGOAL	0.05	0.50	0.51	0.95 (2.07)	0.02	0.24	0.24	0.95 (0.96)	0.02	0.46	0.46	0.95 (1.89)
$n = 500$												
Targ	-0.00	0.31	0.31	0.95 (1.22)	-0.00	0.13	0.13	0.96 (0.55)	-0.00	0.28	0.28	0.97 (1.08)
Targn	-0.00	0.31	0.31	0.95 (1.22)	-0.00	0.13	0.13	0.95 (0.53)	-0.00	0.27	0.27	0.96 (1.04)
Conf	-0.01	0.42	0.42	0.95 (1.64)	-0.00	0.13	0.13	0.95 (0.55)	-0.01	0.39	0.39	0.95 (1.54)
PotConf	-0.00	0.47	0.47	0.94 (1.63)	0.01	0.16	0.16	0.94 (0.61)	-0.01	0.42	0.42	0.94 (1.43)
All	0.00	0.50	0.50	0.96 (1.98)	0.01	0.16	0.16	0.96 (0.68)	-0.01	0.45	0.45	0.97 (1.76)
OAL	0.04	0.32	0.33	0.95 (1.28)	0.06	0.15	0.16	0.94 (0.60)	-0.02	0.26	0.26	0.96 (1.07)
MOAL	0.04	0.32	0.33	0.95 (1.28)	0.00	0.14	0.14	0.96 (0.58)	0.03	0.28	0.29	0.96 (1.16)
GOAL	0.02	0.33	0.33	0.95 (1.29)	0.05	0.15	0.16	0.94 (0.61)	-0.03	0.26	0.26	0.96 (1.07)
MGOAL	0.02	0.33	0.33	0.95 (1.29)	-0.00	0.14	0.14	0.96 (0.59)	0.02	0.29	0.29	0.96 (1.17)
$n = 1000$												
Targ	0.01	0.21	0.22	0.96 (0.85)	0.00	0.10	0.10	0.96 (0.38)	0.01	0.19	0.19	0.94 (0.75)
Targn	0.01	0.21	0.22	0.96 (0.85)	0.00	0.09	0.09	0.96 (0.37)	0.01	0.19	0.19	0.94 (0.73)
Conf	0.01	0.29	0.29	0.95 (1.16)	0.00	0.10	0.10	0.96 (0.39)	0.01	0.28	0.28	0.95 (1.09)
PotConf	0.01	0.31	0.31	0.95 (1.17)	-0.00	0.11	0.11	0.95 (0.44)	0.01	0.28	0.28	0.94 (1.02)
All	0.01	0.32	0.32	0.96 (1.26)	-0.00	0.11	0.11	0.96 (0.46)	0.01	0.29	0.29	0.96 (1.10)
OAL	0.03	0.23	0.23	0.96 (0.93)	0.03	0.11	0.12	0.94 (0.46)	-0.00	0.19	0.19	0.95 (0.76)
MOAL	0.03	0.23	0.23	0.96 (0.93)	0.00	0.10	0.10	0.97 (0.40)	0.03	0.21	0.21	0.96 (0.83)
GOAL	0.02	0.23	0.23	0.96 (0.94)	0.03	0.11	0.11	0.95 (0.46)	-0.01	0.19	0.19	0.95 (0.77)
MGOAL	0.02	0.23	0.23	0.96 (0.94)	-0.00	0.10	0.10	0.97 (0.41)	0.02	0.21	0.21	0.95 (0.84)

Tableau 3.5 IPTW estimates for the average total effect (ATE), average controlled direct effect (ACDE) and average indirect effect (AIE) along with mean bootstrap estimate (Mean), bootstrap standard deviation (Sd) and both 95% bootstrap normal confidence interval (95% NCI) and 95% percentile confidence intervals (95% PCI) for All, OAL, MOAL, GOAL and MGOAL using $n = 918$ observations from the Harvard School of Public Health College Alcohol Study in 1999 (results based on 10 000 bootstrap resamples). The exposure is marijuana use in past 12 months, the mediator is school disengagement since the beginning of the school year and the outcome is the current year grade point average (GPA).

Effect		**Estimate**	**Mean**	**Sd**	**95% Normal CI**	**95% Percentile CI**
	All	-0.115	-0.118	0.050	-0.213 to -0.018	-0.205 to -0.011
ATE	OAL	-0.078	-0.091	0.047	-0.171 to 0.014	-0.188 to 0.012
	MOAL	-0.078	-0.091	0.047	-0.171 to 0.014	-0.188 to 0.012
	GOAL	-0.079	-0.093	0.048	-0.172 to 0.015	-0.192 to 0.010
	MGOAL	-0.079	-0.093	0.048	-0.172 to 0.015	-0.192 to 0.010
	All	-0.050	-0.053	0.050	-0.148 to 0.047	-0.160 to 0.048
	OAL	-0.004	-0.021	0.047	-0.095 to 0.088	-0.128 to 0.075
ACDE	MOAL	-0.028	-0.026	0.046	-0.119 to 0.062	-0.123 to 0.063
	GOAL	-0.004	-0.023	0.048	-0.097 to 0.089	-0.137 to 0.072
	MGOAL	-0.031	-0.029	0.047	-0.123 to 0.062	-0.140 to 0.060
	All	-0.065	-0.065	0.015	-0.094 to -0.036	-0.093 to -0.038
	OAL	-0.075	-0.070	0.015	-0.105 to -0.045	-0.099 to -0.042
AIE	MOAL	-0.050	-0.065	0.018	-0.086 to -0.014	-0.099 to -0.029
	GOAL	-0.074	-0.069	0.015	-0.104 to -0.045	-0.098 to -0.041
	MGOAL	-0.048	-0.064	0.020	-0.087 to -0.009	-0.099 to -0.018

Tableau 3.6 Covariate selection percentage (%) for ATE and ACDE using MOAL and MGOAL for 10 000 bootstrap resamples. The exposure is marijuana use in past 12 months, the mediator is school disengagement since the beginning of the school year and the outcome is the current year grade point average (GPA)

	ATE		**ACDE**	
	MOAL	**MGOAL**	**MOAL**	**MGOAL**
Gender (male :1 / female: 0)	100	100	97	97
Race (White : 1 / Other: 0)	90	89	90	90
Smoke cigarette (yes : 1 / no: 0)	65	65	78	78
Education satisfaction (yes : 1 / no: 0)	59	58	47	47
Parents education (yes : 1 / no: 0)	58	57	62	61
Alcohol drinking and binge (yes : 1 / no: 0)	42	41	22	22
Member of fraternity/sorority (yes : 1 / no: 0)	33	33	38	37
Parents approval of drinking (yes : 1 / no: 0)	31	32	37	36
Number of other drug use	25	24	22	21
High school athletics (yes : 1 / no: 0)	18	19	18	18

Figure 3.5 Probability of covariate being included in PS model for estimating the average total effect (ATE) using OAL/MOAL and GOAL/MGOAL under Scenarios 1-3 (by row) with sample sizes $n = 200, 500, 1000$, based on adaptive weights computed using the unpenalized coefficient estimates obtained by fitting a linear regression of Y on A, M and X. Covariate index: 1, C_{AY}; 2, C_{MY}; 3, C_{AM}; 4, P_Y; 5, P_A; 6, P_M ; 7-20, spurious.

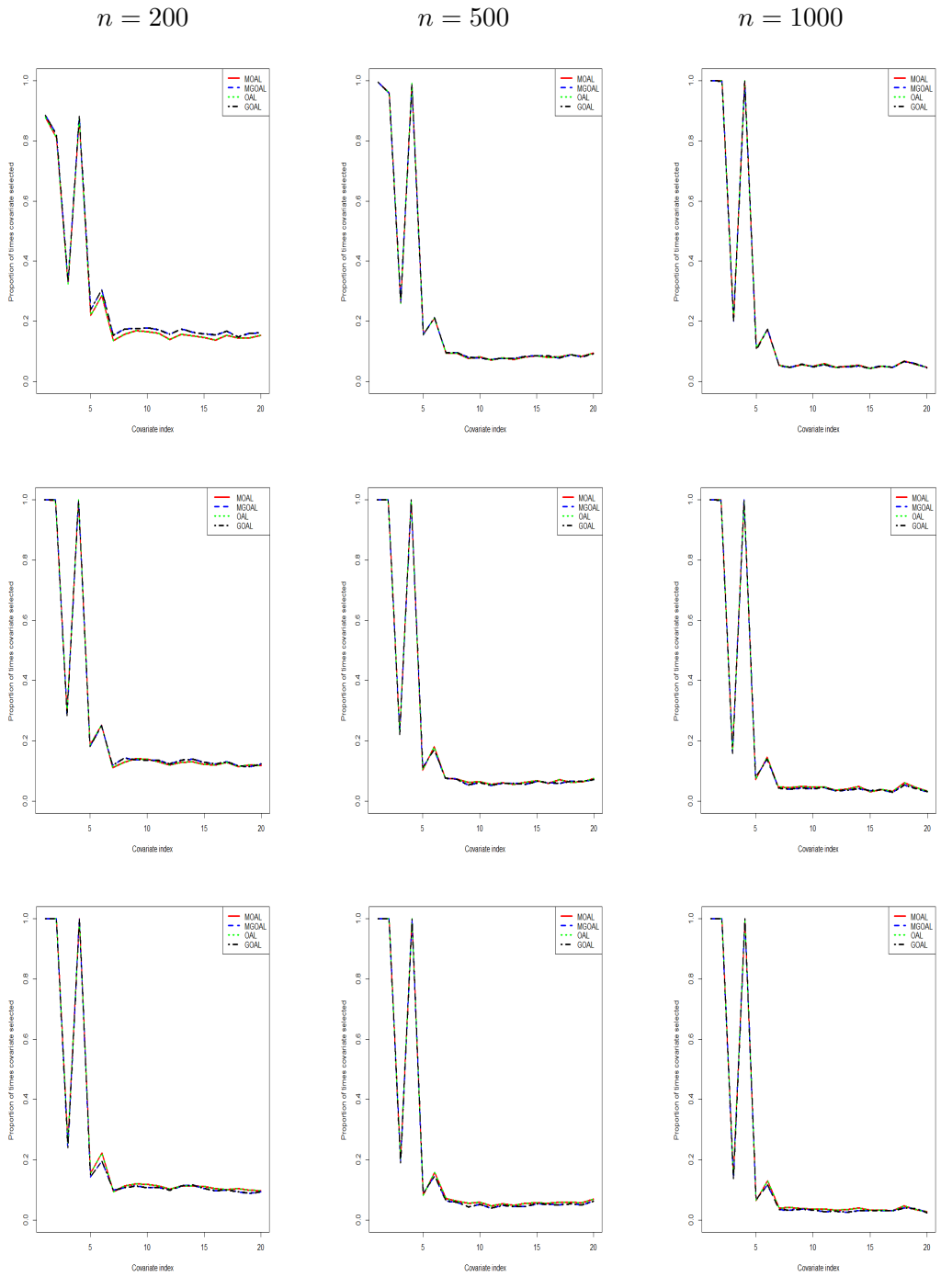


Tableau 3.7 Bias, standard error (SE) and root mean squared error (RMSE) of inverse probability treatment weighting estimators based on OAL, MOAL, GOAL, MGOAL for the average total effect (ATE), average controlled direct effect (ACDE) and average indirect effect (AIE) under Scenario 1 (with adaptive weights computed using the covariates' unpenalized coefficient estimates obtained by fitting a linear regression of Y on A, M and X), with sample sizes $n = 200, 500, 1000$ (results based on 1000 estimates).

	Total Effect			**Direct Effect**			**Indirect Effect**		
	Bias	**SE**	**RMSE**	**Bias**	**SE**	**RMSE**	**Bias**	**SE**	**RMSE**
$n = 200$									
OAL	0.54	0.64	0.84	0.02	0.19	0.19	0.52	0.61	0.80
MOAL	0.54	0.64	0.84	0.02	0.19	0.19	0.52	0.61	0.80
GOAL	0.53	0.65	0.84	0.01	0.19	0.19	0.52	0.62	0.81
MGOAL	0.53	0.65	0.84	0.01	0.19	0.19	0.52	0.62	0.81
$n = 500$									
OAL	0.57	0.41	0.70	0.00	0.11	0.11	0.57	0.40	0.69
MOAL	0.57	0.41	0.70	0.00	0.11	0.11	0.57	0.40	0.69
GOAL	0.57	0.42	0.70	-0.00	0.12	0.12	0.57	0.41	0.70
MGOAL	0.57	0.42	0.70	-0.00	0.12	0.12	0.57	0.41	0.70
$n = 1000$									
OAL	0.60	0.31	0.68	0.00	0.08	0.08	0.60	0.30	0.67
MOAL	0.60	0.31	0.68	0.00	0.08	0.08	0.60	0.30	0.67
GOAL	0.60	0.31	0.68	-0.00	0.08	0.08	0.60	0.31	0.68
MGOAL	0.60	0.31	0.68	-0.00	0.08	0.08	0.60	0.31	0.68

Tableau 3.8 Bias, standard error (SE) and root mean squared error (RMSE) of inverse probability treatment weighting estimators based on OAL, MOAL, GOAL, MGOAL for the average total effect (ATE), average controlled direct effect (ACDE) and average indirect effect (AIE) under Scenario 2 (with adaptive weights computed using the covariates' unpenalized coefficient estimates obtained by fitting a linear regression of Y on A, M and X), with sample sizes $n = 200, 500, 1000$ (results based on 1000 estimates).

	Total Effect			**Direct Effect**			**Indirect Effect**		
	Bias	**SE**	**RMSE**	**Bias**	**SE**	**RMSE**	**Bias**	**SE**	**RMSE**
$n = 200$									
OAL	0.56	0.62	0.84	0.02	0.19	0.19	0.54	0.59	0.80
MOAL	0.56	0.62	0.84	0.02	0.19	0.19	0.54	0.59	0.80
GOAL	0.55	0.64	0.84	0.00	0.19	0.19	0.54	0.60	0.81
MGOAL	0.55	0.64	0.84	0.00	0.19	0.19	0.54	0.60	0.81
$n = 500$									
OAL	0.58	0.40	0.70	0.01	0.12	0.12	0.57	0.39	0.69
MOAL	0.58	0.40	0.70	0.01	0.12	0.12	0.57	0.39	0.69
GOAL	0.57	0.41	0.70	-0.00	0.12	0.12	0.57	0.40	0.70
MGOAL	0.57	0.41	0.70	-0.00	0.12	0.12	0.57	0.40	0.70
$n = 1000$									
OAL	0.61	0.30	0.68	0.01	0.09	0.09	0.60	0.30	0.67
MOAL	0.61	0.30	0.68	0.01	0.09	0.09	0.60	0.30	0.67
GOAL	0.60	0.31	0.68	-0.00	0.09	0.09	0.60	0.30	0.68
MGOAL	0.60	0.31	0.68	-0.00	0.09	0.09	0.60	0.30	0.68

Tableau 3.9 Bias, standard error (SE) and root mean squared error (RMSE) of inverse probability treatment weighting estimators based on OAL, MOAL, GOAL, MGOAL for the average total effect (ATE), average controlled direct effect (ACDE) and average indirect effect (AIE) under Scenario 3 (with adaptive weights computed using the covariates' unpenalized coefficient estimates obtained by fitting a linear regression of Y on A, M and X), with sample sizes $n = 200, 500, 1000$ (results based on 1000 estimates).

	Total Effect			**Direct Effect**			**Indirect Effect**		
	Bias	**SE**	**RMSE**	**Bias**	**SE**	**RMSE**	**Bias**	**SE**	**RMSE**
$n = 200$									
OAL	0.59	0.62	0.85	0.03	0.20	0.20	0.56	0.58	0.80
MOAL	0.59	0.62	0.85	0.03	0.20	0.20	0.56	0.58	0.80
GOAL	0.57	0.63	0.85	0.00	0.20	0.20	0.56	0.60	0.82
MGOAL	0.57	0.63	0.85	0.00	0.20	0.20	0.56	0.60	0.82
$n = 500$									
OAL	0.59	0.39	0.71	0.01	0.12	0.12	0.58	0.38	0.70
MOAL	0.59	0.39	0.71	0.01	0.12	0.12	0.58	0.38	0.70
GOAL	0.58	0.40	0.71	-0.00	0.12	0.12	0.58	0.39	0.70
MGOAL	0.58	0.40	0.71	-0.00	0.12	0.12	0.58	0.39	0.70
$n = 1000$									
OAL	0.62	0.30	0.69	0.01	0.09	0.09	0.61	0.29	0.68
MOAL	0.62	0.30	0.69	0.01	0.09	0.09	0.61	0.29	0.68
GOAL	0.61	0.30	0.68	-0.00	0.09	0.09	0.61	0.29	0.68
MGOAL	0.61	0.30	0.68	-0.00	0.09	0.09	0.61	0.29	0.68

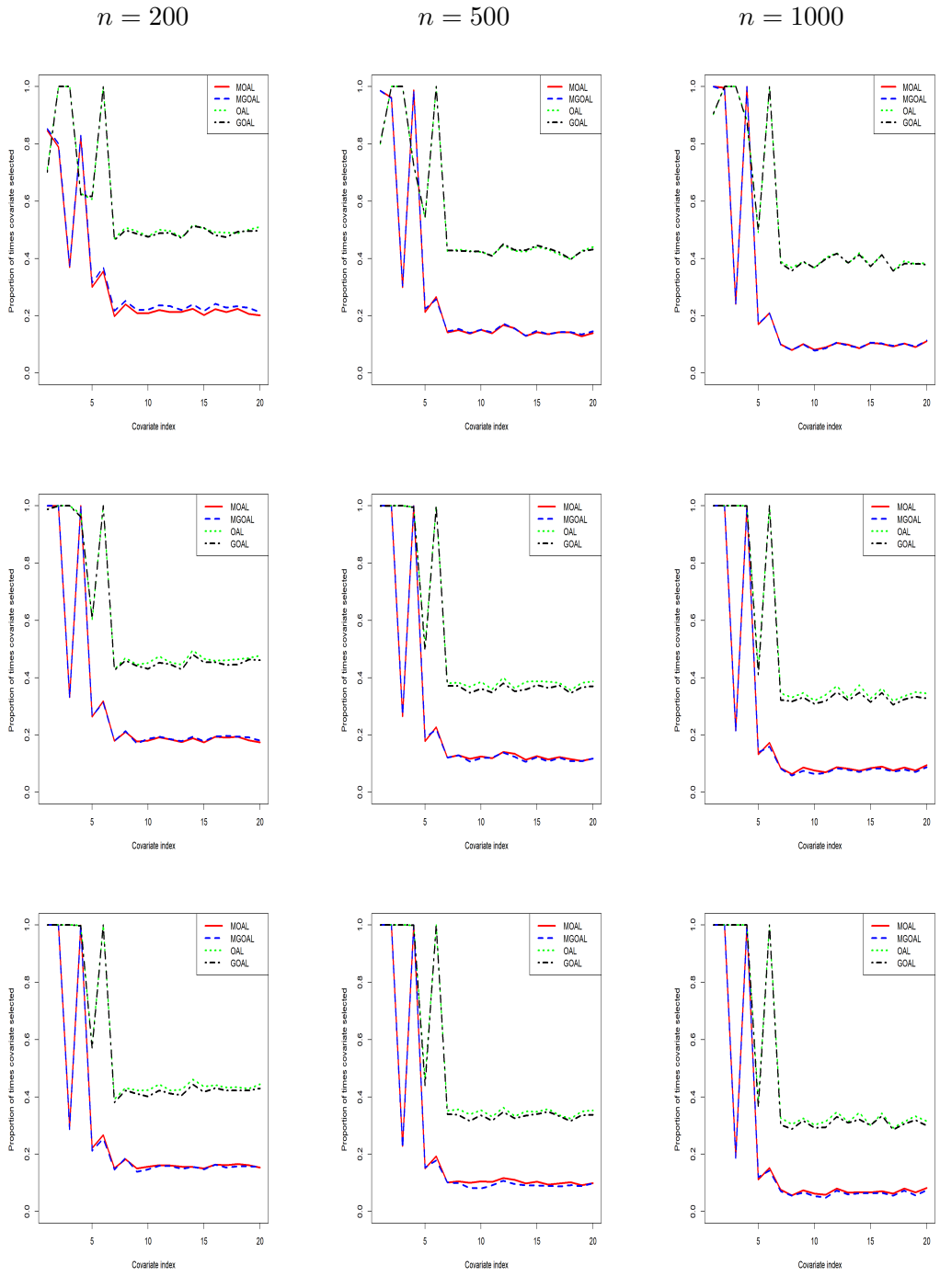
Tableau 3.10 Bias, standard error (SE) and root mean squared error (RMSE) and coverage probability (confidence interval length) of the nine studied inverse probability treatment weighting estimators (Targ, Targn, Conf, PotConf, ALL, OAL, MOAL, GOAL, MGOAL) for the average total effect (ATE), average controlled direct effect (ACDE) and average indirect effect (AIE) under Scenario 2 (in presence of intermediate confounding), with sample sizes $n = 200, 500, 1000$ (results based on 1000 estimates).

	Total Effect				**Direct Effect**				**Indirect Effect**			
	Bias	**SE**	**RMSE**	**P (Len)**	**Bias**	**SE**	**RMSE**	**P (Len)**	**Bias**	**SE**	**RMSE**	**P (Len)**
Targ	0.02	0.52	0.52	0.96 (2.08)	0.02	0.24	0.24	0.95 (0.92)	0.00	0.46	0.46	0.96 (1.86)
Targn	0.02	0.52	0.52	0.96 (2.08)	0.02	0.24	0.24	0.95 (0.90)	0.00	0.44	0.43	0.96 (1.76)
Conf	0.04	0.71	0.71	0.94 (2.77)	0.02	0.25	0.25	0.95 (0.98)	0.02	0.66	0.66	0.94 (2.60)
PotConf	0.05	0.74	0.74	0.96 (2.79)	0.03	0.31	0.31	0.94 (1.10)	0.02	0.62	0.62	0.96 (2.38)
All	0.04	0.93	0.93	1.00 (4.79)	0.03	0.36	0.36	0.98 (1.65)	0.01	0.78	0.78	1.00 (4.12)
OAL	0.10	0.52	0.53	0.96 (2.16)	0.09	0.27	0.28	0.95 (1.10)	0.01	0.43	0.43	0.95 (1.76)
MOAL	0.10	0.52	0.53	0.96 (2.16)	0.03	0.24	0.24	0.95 (0.96)	0.07	0.47	0.47	0.96 (1.98)
GOAL	0.06	0.53	0.53	0.96 (2.20)	0.08	0.28	0.29	0.95 (1.12)	-0.01	0.43	0.43	0.96(1.78)
MGOAL	0.06	0.53	0.53	0.96 (2.20)	0.01	0.24	0.24	0.95 (0.98)	0.05	0.48	0.48	0.96 (2.01)
Targ	-0.00	0.32	0.32	0.95 (1.24)	0.00	0.14	0.14	0.96 (0.56)	-0.00	0.28	0.28	0.96 (1.09)
Targn	-0.00	0.32	0.32	0.95 (1.24)	0.00	0.13	0.13	0.95 (0.55)	-0.00	0.27	0.27	0.96 (1.04)
Conf	-0.01	0.45	0.45	0.95 (1.75)	0.00	0.15	0.15	0.96 (0.60)	-0.01	0.41	0.41	0.95 (1.64)
PotConf	-0.00	0.49	0.49	0.94 (1.69)	0.01	0.17	0.17	0.95 (0.67)	-0.01	0.42	0.42	0.94 (1.43)
ALL	0.00	0.53	0.53	0.96 (2.07)	0.01	0.19	0.19	0.96 (0.77)	-0.01	0.45	0.45	0.98 (1.76)
OAL	0.02	0.33	0.33	0.96 (1.35)	0.02	0.14	0.15	0.97 (0.66)	0.00	0.27	0.27	0.97 (1.11)
MOAL	0.02	0.33	0.33	0.96 (1.35)	0.01	0.14	0.14	0.96 (0.58)	0.02	0.29	0.29	0.97 (1.22)
GOAL	0.00	0.33	0.33	0.97 (1.36)	0.01	0.14	0.14	0.97 (0.67)	-0.01	0.27	0.27	0.97 (1.11)
MGOAL	0.00	0.33	0.33	0.97 (1.36)	-0.00	0.13	0.13	0.96 (0.59)	0.01	0.29	0.29	0.97 (1.22)
Targ	0.01	0.22	0.22	0.95 (0.87)	0.00	0.10	0.10	0.95 (0.39)	0.01	0.20	0.20	0.94 (0.76)
Targn	0.01	0.22	0.22	0.95 (0.87)	0.00	0.10	0.10	0.95 (0.38)	0.01	0.19	0.19	0.94 (0.73)
Conf	0.01	0.31	0.31	0.95 (1.23)	0.00	0.11	0.11	0.95 (0.42)	0.01	0.30	0.30	0.95 (1.15)
PotConf	0.01	0.33	0.33	0.95 (1.21)	-0.00	0.12	0.12	0.96 (0.48)	0.01	0.28	0.28	0.94 (1.02)
ALL	0.01	0.34	0.34	0.96 (1.31)	-0.00	0.13	0.13	0.96 (0.51)	0.01	0.29	0.29	0.96 (1.10)
OAL	0.02	0.23	0.23	0.96 (0.94)	0.01	0.10	0.10	0.96 (0.42)	0.01	0.20	0.20	0.96 (0.78)
MOAL	0.02	0.23	0.23	0.96 (0.94)	0.01	0.10	0.10	0.96 (0.41)	0.02	0.21	0.21	0.96 (0.84)
GOAL	0.01	0.23	0.23	0.96 (0.94)	0.01	0.10	0.10	0.96 (0.43)	-0.00	0.20	0.20	0.95 (0.78)
MGOAL	0.01	0.23	0.23	0.96 (0.94)	-0.00	0.10	0.10	0.97 (0.41)	0.01	0.21	0.21	0.96 (0.84)

Tableau 3.11 Bias, standard error (SE), root mean squared error (RMSE) and coverage probability (confidence interval length) of the nine studied inverse probability treatment weighting estimators (Targ, Targn, Conf, PotConf, ALL, OAL, MOAL, GOAL, MGOAL) for the average total effect (ATE), average controlled direct effect (ACDE) and average indirect effect (AIE) under Scenario 3 (in presence of intermediate confounding), with sample sizes $n = 200, 500, 1000$ (results based on 1000 estimates).

	Total Effect				**Direct Effect**				**Indirect Effect**			
	Bias	**SE**	**RMSE**	**P (Len)**	**Bias**	**SE**	**RMSE**	**P (Len)**	**Bias**	**SE**	**RMSE**	**P (Len)**
Targ	0.03	0.53	0.54	0.97 (2.15)	0.02	0.25	0.25	0.96 (0.97)	0.01	0.47	0.47	0.96 (1.90)
Targn	0.03	0.53	0.54	0.97 (2.15)	0.02	0.25	0.25	0.96 (0.97)	0.00	0.44	0.43	0.96 (1.76)
Conf	0.05	0.76	0.76	0.94 (2.98)	0.02	0.29	0.29	0.94 (1.12)	0.03	0.70	0.70	0.94 (2.77)
PotConf	0.06	0.78	0.78	0.96 (2.92)	0.04	0.36	0.36	0.95 (1.27)	0.02	0.62	0.62	0.96 (2.38)
ALL	0.05	0.98	0.98	1.00 (5.10)	0.04	0.43	0.43	0.99 (2.10)	0.01	0.78	0.78	1.00 (4.12)
OAL	0.10	0.53	0.53	0.96 (2.23)	0.07	0.26	0.27	0.97 (1.13)	0.03	0.43	0.43	0.95 (1.80)
MOAL	0.10	0.53	0.53	0.96 (2.23)	0.04	0.24	0.24	0.96 (1.00)	0.06	0.47	0.48	0.96 (2.04)
GOAL	0.06	0.53	0.53	0.96 (2.27)	0.05	0.26	0.27	0.97 (1.16)	0.01	0.43	0.43	0.96 (1.82)
MGOAL	0.06	0.53	0.53	0.96 (2.27)	0.01	0.24	0.24	0.96 (1.01)	0.05	0.48	0.48	0.96 (2.08)
Targ	0.00	0.33	0.33	0.95 (1.27)	0.00	0.14	0.14	0.96 (0.58)	-0.00	0.29	0.29	0.96 (1.10)
Targn	0.00	0.33	0.33	0.95 (1.27)	0.00	0.14	0.14	0.96 (0.58)	-0.00	0.27	0.27	0.96 (1.04)
Conf	-0.01	0.48	0.48	0.96 (1.88)	0.00	0.17	0.17	0.96 (0.69)	-0.01	0.44	0.44	0.95 (1.74)
PotConf	0.00	0.52	0.52	0.94 (1.77)	0.01	0.21	0.21	0.94 (0.77)	-0.01	0.42	0.42	0.94 (1.43)
ALL	0.00	0.57	0.57	0.97 (2.20)	0.01	0.23	0.23	0.96 (0.94)	-0.01	0.45	0.45	0.98 (1.76)
OAL	0.02	0.34	0.34	0.96 (1.36)	0.02	0.15	0.15	0.96 (0.64)	-0.00	0.28	0.28	0.96 (1.11)
MOAL	0.02	0.34	0.34	0.96 (1.36)	0.01	0.14	0.14	0.96 (0.60)	0.01	0.30	0.30	0.97 (1.23)
GOAL	-0.00	0.34	0.34	0.96 (1.37)	0.01	0.15	0.15	0.97 (0.64)	-0.01	0.28	0.28	0.96 (1.12)
MGOAL	-0.00	0.34	0.34	0.96 (1.37)	-0.00	0.14	0.14	0.96 (0.60)	0.00	0.30	0.30	0.98 (1.23)
Targ	0.01	0.22	0.22	0.95 (0.88)	0.00	0.10	0.10	0.95 (0.41)	0.01	0.20	0.20	0.94 (0.77)
Targn	0.01	0.22	0.22	0.95 (0.88)	0.01	0.10	0.10	0.95 (0.40)	0.01	0.19	0.19	0.94 (0.73)
Conf	0.01	0.34	0.34	0.95 (1.32)	0.00	0.13	0.13	0.95 (0.49)	0.01	0.31	0.31	0.95 (1.23)
PotConf	0.01	0.35	0.35	0.94 (1.27)	-0.00	0.15	0.15	0.95 (0.56)	0.01	0.28	0.28	0.94 (1.02)
ALL	0.01	0.36	0.36	0.96 (1.38)	-0.00	0.16	0.16	0.96 (0.61)	0.01	0.29	0.29	0.96 (1.10)
OAL	0.02	0.23	0.23	0.96 (0.96)	0.01	0.10	0.10	0.96 (0.44)	0.01	0.19	0.19	0.95 (0.78)
MOAL	0.02	0.23	0.23	0.96 (0.96)	0.01	0.10	0.10	0.96 (0.42)	0.02	0.21	0.21	0.96 (0.85)
GOAL	0.00	0.23	0.23	0.96 (0.95)	0.00	0.10	0.10	0.96 (0.44)	-0.00	0.20	0.20	0.96 (0.78)
MGOAL	0.00	0.23	0.23	0.96 (0.95)	-0.00	0.10	0.10	0.97 (0.42)	0.01	0.21	0.21	0.96 (0.85)

Figure 3.6 Probability of covariate being included in PS model for estimating the average controlled direct effect (ACDE) under Scenarios 1-3 (by row) with sample sizes $n = 200, 500, 1000$, in presence of intermediate confounding. Covariate index: 1, C_{AY}; 2, C_{MY}; 3, C_{AM}; 4, P_Y; 5, P_A; 6, P_M ; 7-20, spurious.



CHAPITRE IV

TROISIÈME ARTICLE: GRAPHICAL LASSO NETWORKS IN THE CONTEXT OF CAUSAL MEDIATION ANALYSIS

Ismaila Baldé and Geneviève Lefebvre

Le Chapitre 4 est constitué d'un article original dont je suis le premier auteur. Cet article est en préparation et sera accompagné d'une application Shiny `R` qui est en construction. Dans cet article, j'ai collaboré avec ma directrice de thèse, Geneviève Lefebvre, professeure au Département de mathématiques à l'UQAM. J'ai développé l'idée originale, effectué la revue de la littérature, réalisé les analyses statistiques et rédigé le code informatique et le manuscrit.

Abstract: Graphical lasso (glasso) is a very popular and fast regularization approach developed by Friedman, Hastie and Tibshirani (2008) to estimate the inverse of the covariance matrix from observed multivariate normal data. This inverse covariance matrix is interesting since it encodes the conditional associations between the variables, which can be useful to understand the relationships between them. In this paper, we study the value of the graphical lasso in the context of causal mediation analysis and propose an adjustment to better take into account the particularities of the targeted analysis context. A shiny application, currently in progress, aims to offer a simple graphical interface that allows users to have a tool for visualizing direct relationships and probable mediation between variables.

Keywords: Observational data, Graphical lasso, Causal mediation analysis, Shiny application, Variable selection.

4.1 Introduction

Graphical models are powerful statistical tools that organize variables into a graphical network displaying the dependencies between a potentially very large set of covariates (Strobl *et al.*, 2012). Graphical models for health applications use such data types as, for example, demographic, clinical, phenotypic, biometric and multi-omic variables. These models can provide a data-rich picture of complex health processes, such as comorbidity, by learning connected clusters of variables, and can thus enrich traditional descriptive statistics and regression models (Zierer *et al.*, 2016; Kalisch *et al.*, 2010).

Graphical models were proposed as a way to reveal conditional dependencies between pairs of variables while eliminating mediated associations (Zierer *et al.*, 2016). This model is composed of a vertex set corresponding to the variables of interest where every vertex pair can be connected by an edge representing the conditional dependency (correlation) between the two variables, that is, the dependency between two variables once we control for all other variables. In a fully mediated context with no direct effect, there would be no edge between treatment A and outcome Y (since $A \amalg Y|M$), while each of them would be connected to the mediator M. Although the direction of the associations and, thus, causality cannot in general be determined by these models, the

resulting network of associations is nevertheless thought to be a useful exploratory tool. The classical model for continuous data assumes that the observations are independent and identically distributed from a multivariate Gaussian distribution with mean μ and covariance matrix $\mathbf{\Sigma}$, that is $X = (X_1, \ldots, X_p) \sim N(\mu, \mathbf{\Sigma})$. The inverse covariance matrix $\mathbf{\Theta} = \mathbf{\Sigma^{-1}}$ (precision or concentration matrix), contains all the conditional dependence information for the multivariate Gaussian model. In particular, the zero entries of the inverse covariance matrix $\Theta_{ij} = 0$ correspond to conditionally independent pairs of variables X_i and X_j (that is, $X_i \amalg X_j | X_{-\{i,j\}}$ where $X_{-\{i,j\}}$ represents all variables except X_i and X_j).

Sparse graphical networks are usually desired. To induce sparsity in the number of edges detected, methodologies based on lasso regularization, such as the well-known graphical lasso (glasso) (Friedman *et al.*, 2008) can be considered. In the last two decades, the estimation of sparse undirected graphical models through the use of lasso regularization or its variants has attracted much attention in various statistical applications such as genetic networks, social networks and financial markets (Uhler, 2017).

A first objective of this article is to investigate whether the graphical lasso can be effective to understand the role of covariates in causal mediation. In particular, we are interested in assessing whether graphical lasso can help retrieve covariate types (e.g., treatment-outcome, treatment-mediator, or mediator-outcome confounders, predictors of treatment, mediator or outcome, spurious) to enhance data knowledge. Another contribution of this article is to develop a new graphical lasso algorithm which specifically aims to learn causal mediation skeletons, namely mglasso. A causal mediation skeleton is a causal mediation graph with undirected edges. Our approach is to modify the standard glasso algorithm of Friedman, Hastie and Tibshirani (2008) by using a tailored componentwise penalty instead of a common penalty, and then enjoy the computational advantages of the glasso. We show that our mglasso algorithm is more useful than the standard glasso in the context of causal mediation analysis.

Our paper is organized as follows. Section 2 introduces the overview of the methods approach. Section 2.1 recalls the graphical lasso problem. Section 2.2 introduces the glasso algorithm. Section 2.3 presents our mglasso algorithm. Section 3 describes the

simulation study and results are presented in Section 4. We close the paper with a discussion.

4.2 Methods

4.2.1 Gaussian graphical model

We consider n independent and identically distributed random vectors $X^{(1)}, \ldots, X^{(n)} \sim N(\mu, \mathbf{\Sigma})$. The sample mean ($\bar{X}$) and the sample covariance matrix ($\mathbf{S}$) are defined as

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X^{(i)} \quad \text{and} \quad \mathbf{S} = \frac{1}{n}\sum_{i=1}^{n}\left(X^{(i)} - \bar{X}\right)\left(X^{(i)} - \bar{X}\right)^T.$$

We use the notation and language of Uhler (2017) to describe the undirected gaussian graphical model with a given graph. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an undirected graph with vertices $\mathcal{V} = \{1, 2, \ldots, q\}$ and edges $\mathcal{E}$. We use the notation (i, j) to denote the link between vertices i and j.

A random vector $X \in \mathbb{R}^q$ is said to satisfy the undirected gaussian graphical model with graph $\mathcal{G}$, if $X \sim N_q(\mu, \mathbf{\Sigma})$ with $\mathbf{\Theta}_{ij} = 0$ for all $(i, j) \notin \mathcal{E}$, where $\mathbf{\Theta} = \mathbf{\Sigma}^{-1}$.

Let ℓ_n be the Gaussian log-likelihood of the data; ℓ_n can be expressed as

$$\ell_n(\mu, \Sigma) \propto -\frac{n}{2}\log\det(\mathbf{\Sigma}) - \frac{n}{2}\operatorname{tr}(\mathbf{S}\mathbf{\Sigma}^{-1}) - \frac{n}{2}(\bar{X} - \mu)^T\mathbf{\Sigma}^{-1}(\bar{X} - \mu). \tag{4.1}$$

The reader interested in the proof of Equation (4.1) or others details are referred to the paper of Uhler (2017) and the recent book of Lederer (2022).

The maximum likelihood estimators (MLE) of μ and $\mathbf{\Sigma}$ can be easily deduced from the unconstrained model (where $\mu \in \mathbb{R}^q$ and $\mathbf{\Sigma}$ is a positive semidefinite matrice) in (4.1) $(\mu, \mathbf{\Sigma})$, which are $\hat{\mu} = \bar{X}$ and $\hat{\mathbf{\Sigma}} = \mathbf{S}$, respectively (Friedman *et al.*, 2008; Lederer, 2022; Uhler, 2017).

It is well known that solving Equation (4.1) under general constraints on the parameters $(\mu, \mathbf{\Sigma})$ is very hard (Uhler, 2017). Fortunately, the Gaussian graphical problem restricts the constraints only on the covariance matrix (that is, it assumes μ is unconstrained and $\mathbf{\Theta}$ is constrained to non negative definite matrices) (Uhler, 2017). Hence we let $\hat{\mu} = \bar{X}$ and Equation (4.1) becomes

$$\ell_n(\Sigma) \propto -\frac{n}{2}\log\det(\mathbf{\Sigma}) - \frac{n}{2}\operatorname{tr}(\mathbf{S}\mathbf{\Sigma}^{-1}). \tag{4.2}$$

Let $\mathbf{\Theta} = \mathbf{\Sigma^{-1}}$. Maximizing the problem in Equation (4.2) is equivalent to maximizing the following problem

$$\ell_n(\mu, \Sigma) \propto \log\det \mathbf{\Theta} - \mathrm{tr}(\mathbf{S\Theta}). \tag{4.3}$$

The problem in Equation (4.3) is obtained by ignoring the constant multiplicative factor $\frac{n}{2}$ and the two problems have the same MLE, $\mathbf{\Theta} = \mathbf{S}^{-1}$. That is, the problem in Equation (4.3) is maximized when $\mathbf{\Theta} = \mathbf{S}^{-1}$.

Graphical lasso problem

In general, we would like to avoid estimating spurious edges and encourage sparsity in the estimation of $\mathbf{\Theta}$. One can impose a L_1 penalty for the estimation of $\mathbf{\Theta}$ to increase its sparsity. That is, we can penalize the problem in Equation (4.3) using lasso (Tibshirani, 1996) as

$$\log\det \mathbf{\Theta} - \mathrm{tr}(\mathbf{S\Theta}) - \rho\|\mathbf{\Theta}\|_1, \tag{4.4}$$

where $\|.\|_1$ is the L_1-norm.

Expression (4.4) is the so-called graphical lasso problem. The goal is to maximize (4.4) over non negative definite matrices $\mathbf{\Theta}$.

A number of algorithms have been proposed to solve the graphical lasso problem. Meinshausen and Bühlmann (2006) used a simple technique called lasso neighborhood selection. The neighborhood selection consists in fitting a lasso regression for each variable using the rest as predictors: $\hat{\Theta}_{ij}$ is considered nonzero when either the estimated coefficient of X_i on X_j is nonzero or the estimated coefficient of X_j on X_i is nonzero (Friedman *et al.*, 2008; Meinshausen et Bühlmann, 2006). In addition, Meinshausen and Bühlmann (2006) showed that their method consistently estimates nonzero components of $\mathbf{\Theta}$ for an adequate choice of the regularization parameter. Three different papers established that the simpler algorithm of Meinshausen and Bühlmann (2006) can be viewed as an approximation to the exact graphical problem and proposed algorithms to solve the exact problem based on lasso-penalized likelihood (Banerjee *et al.*, 2008; Yuan et Lin, 2008; Dahl *et al.*, 2008). Friedman, Hastie and Tibshirani (2008) proposed the graphical lasso algorithm (glasso) to maximize the lasso penalized log-likelihood. The glasso is an efficient and very fast algorithm, and is implemented in `R` language package (`glasso`)

(Friedman *et al.*, 2011).

Several variants of glasso have been developed in the last decade. Fan et al. (2017) proposed the L-glasso, a high dimensional semiparametric latent graphical model for mixed data. Meinshausen and Bühlmann (2010) proposed Boglasso, which provides a stable graph estimation and selects edge using frequency thresholding. In this article, we focus on the original glasso for simplicity. In addition to being fast, the glasso is flexible since it allows for different types of penalties. Indeed, one can specify different amounts of penalization for each variable or allow each inverse covariance element to be penalized differently (Friedman *et al.*, 2008).

4.2.2 The glasso (Friedman *et al.*, 2008)

Since Friedman et al. (2008) used the blockwise coordinate descent approach of Banerjee et al. (2008) as a launching point before proposing the glasso algorithm, we briefly recall the Banerjee et al. (2008) method. Banerjee et al. (2008) start by showing that the optimization problem in (4.4) is convex and then solve it by optimizing over each row and corresponding column using block coordinate descent. Let $\mathbf{W}$ be the estimate of $\mathbf{\Sigma}$. Partition $\mathbf{W}$, $\mathbf{S}$, and $\mathbf{\Theta}$ as

$$\mathbf{W} = \begin{pmatrix} \mathbf{W}_{11} & w_{12} \\ w_{12}^T & w_{22} \end{pmatrix} \quad , \quad \mathbf{S} = \begin{pmatrix} \mathbf{S}_{11} & s_{12} \\ s_{12}^T & s_{22} \end{pmatrix} \quad , \quad \mathbf{\Theta} = \begin{pmatrix} \mathbf{\Theta}_{11} & \theta_{12} \\ \theta_{12}^T & \theta_{22} \end{pmatrix}.$$

Banerjee et al. (2008) defined the solution for w_{12} as

$$w_{12} = \arg\min_{v} \{ v^T \mathbf{W}_{11}^{-1} v : \|v - s_{12}\|_\infty \leqslant \rho \}, \tag{4.5}$$

and showed that solving (4.5) is equivalent to solving the dual problem defined as

$$\min_{\beta} \{ \frac{1}{2} \|W_{11}^{1/2} \beta - b\|_2^2 + \rho \|\beta\|_1 \}, \tag{4.6}$$

where $b = \mathbf{W}_{11}^{-1/2} s_{12}$. In other words, if β solves (4.6) then $w_{12} = \mathbf{W}_{11}\beta$ solves (4.5) (Banerjee *et al.*, 2008; Friedman *et al.*, 2008). The problem (4.6) can be viewed as a lasso least-squares problem (Banerjee *et al.*, 2008; Friedman *et al.*, 2008).

Friedman et al. (2008) first verify the equivalence between the solutions of problems (4.4) and (4.6). The subgradient of (4.4) is given by

$$\mathbf{W} - \mathbf{S} - \rho\mathbf{\Gamma} = 0, \tag{4.7}$$

where $\Gamma_{ij} = \text{sign}(\Theta_{ij})$ if $\Theta_{ij} \neq 0$, else $\Gamma_{ij} \in [-1, 1]$ if $\Theta_{ij} = 0$. Let $\nu \in \text{sign}(\beta)$. Now we can compare the upper right block of (4.7)

$$w_{12} - s_{12} - \rho\gamma_{12} = 0 \tag{4.8}$$

and the subgradient of (4.6)

$$\mathbf{W}_{11}\beta - s_{12} + \rho\nu = 0. \tag{4.9}$$

The solutions of (4.7) and (4.9) are equivalent if

$$\beta = \mathbf{W}_{11}^{-1} w_{12} \quad \text{and} \quad \nu = -\gamma_{12}.$$

Finally, the problems (4.4) and (4.6) are equivalent. From $\mathbf{W\Theta} = \mathbf{I}$, we have $\theta_{12} = -\theta_{22}\beta$. Thus, the solution of (4.6) gives (up to a negative constant) the corresponding part of $\mathbf{\Theta}$ ($\theta_{12} = -\theta_{22}\beta$). If $\mathbf{W}_{11} = \mathbf{S}_{11}$, the solution $\hat{\beta}$ is easily seen to equal the solution proposed by Meinshausen and Bühlmann (2006), however $\mathbf{W}_{11} \neq \mathbf{S}_{11}$ in general. Thus, Meinshausen and Bühlmann's approach (Meinshausen et Bühlmann, 2006) does not yield the exact solution of (4.4) (Friedman *et al.*, 2008). In the sequel we present the graphical lasso algorithm (glasso) of Friedman et al. (2008).

Algorithm ***glasso*** (Friedman *et al.*, 2008)

1: Start with $\mathbf{W} = \mathbf{S} + \rho\mathbf{I}$;

2: Solve the lasso problem (4.6) and save the value of $\hat{\boldsymbol{\beta}}$;

3: Update w_{12} and w_{21} using the relation $w_{12} = \mathbf{W}_{11}\hat{\beta}$;

4: Rearrange $\mathbf{W}$ so the next row and column are in position $i = 1, j = 2$;

5: Repeat steps 2-4 until convergence ;

6: Calculate the diagonal of $\boldsymbol{\Theta}$ using $\hat{\theta}_{22} = 1/(w_{22} - w_{12}^T\hat{\beta})$;

7: Use the most recent value of β to complete $\boldsymbol{\Theta}$ (using $\beta = -\frac{\theta_{12}}{\theta_{22}}$).

4.2.3 The proposed mglasso

Friedman et al. (2008) solved the problem (4.4) via glasso based using the penalty $\rho\|\boldsymbol{\Theta}\|_1$. However, they showed that the glasso algorithm is flexible and can be used in different situations with different penalties. They showed that one can consider a matrix penalty, namely $\|\boldsymbol{\Theta} * \mathbf{P}_\rho\|_1$. The matrix $\mathbf{P}_\rho$ is defined as $\mathbf{P}_\rho = \{\rho_{jk}\}$, where $\rho_{jk} = \rho_{kj}$ and $*$ is the componentwise multiplication. In this case, Friedman et al. (2008) defined the graphical lasso problem as the maximization over $\boldsymbol{\Theta}$ of

$$\log\det\boldsymbol{\Theta} - \mathrm{tr}(\mathbf{S}\boldsymbol{\Theta}) - \|\boldsymbol{\Theta} * \mathbf{P}_\rho\|_1. \tag{4.10}$$

Expression (4.10) can be maximized by the preceeding glasso algorithm by replacing ρ with ρ_{jk} (Friedman *et al.*, 2008).

Consider the causal mediation directed acyclic graph (DAG) in Figure 4.1, displaying the causal relationships between the variables.

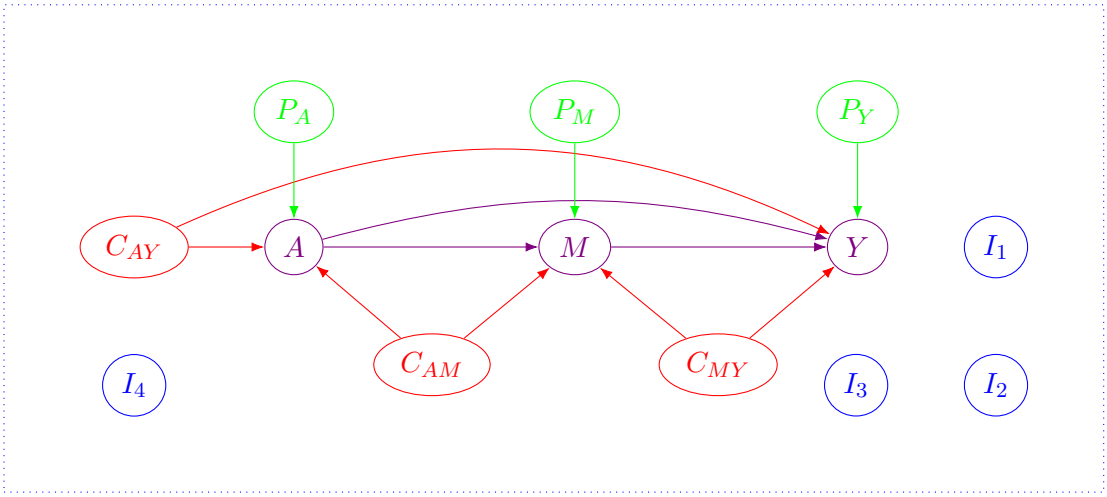


Figure 4.1 Hypothesized causal mediation graph.

The variables in Figure 4.1 are the covariates $L = (C_{AY}, C_{MY}, C_{AM}, P_Y, P_A, P_M, I_S)$, exposure (A), mediator (M) and outcome (Y). The baseline covariates (L) include exposure-outcome confounders (C_{AY}), mediator-outcome confounders (C_{MY}), exposure-mediator confounders (C_{AM}), pure predictors of outcome (P_Y), pure predictors of exposure (P_A), pure predictors of mediator (P_M) and spurious covariates $(I_S = (I_1, I_2, I_3, I_4))$. Considering a mediation analysis context, we would like to assess the presence of edges between input variables X and, for each variable except the posited treatment, mediator and outcome (that is for $X_{-\{A,M,Y\}}$), determine its type according to the possible labels $(C_{AY}, C_{MY}, C_{AM}, P_Y, P_A, P_M, I_S)$. We focus on maximizing the Expression (4.10) over Θ to help achieve this objective.

It is important to note that the network of conditional dependencies induced from the DAG in Figure 4.1 does not correspond to the skeleton of this DAG, but rather to its moral graph. The skeleton of a DAG is the undirected graph obtained by replacing the directed edges with undirected edges. As such, the skeleton contains an edge between two nodes if and only if the underlying DAG contains an edge between those nodes (Spirtes *et al.*, 2000). Put differently, the skeleton informs on which variables do and do not share a link, but does not inform on the directionality of the link. The moral

graph of a DAG is an undirected graph that is obtained by first connecting all pairs of non-adjacent nodes (that is, nodes not connected directly by an edge) that have a common child (direct descendant) in the DAG, and then replacing all directed edges with undirected edges. The moral graph thus features an undirected edge between two nodes if these are either connected by a directed edge in the DAG or these are involved in a collider fork. In general, the moral graph of a DAG contains more edges than the skeleton (Ryan *et al.*, 2022).

While the moral graph of a causal mediation graph can be useful, we argue that humans are better able to interpret skeleton graphs. In fact, it is not infrequent that applied researchers erroneously interpret a conditional dependency network as the skeleton of a causal graph, ignoring the fact that some edges can be due to conditioning on colliders (Ryan *et al.*, 2022). With this in mind, we introduce our mglasso algorithm, which departs from the glasso mainly based on the specification of the $\mathbf{P}_\rho$ matrix based on adaptive weights. Shortreed and Ertefaie (2017) and Baldé et al. (2022) showed that using a modified adaptive lasso or adaptive elastic net leads to better results in variable selection for causal inference. These two papers constructed their adaptive weights from the outcome model. We adopt a similar strategy in our modified glasso algorithm for mediation, named mglasso. First, we let $X_L = X_{-\{A,M,Y\}}$; this notation is used to indicate that we aim to classify each variable in $X_{-\{A,M,Y\}}$ as one covariate type in L, that is either $C_{AY}, C_{MY}, C_{AM}, P_Y, P_A, P_M$ or I_S. To construct the matrix $\mathbf{P}_\rho$, the weights are taken inversely proportional to the regression coefficients of the outcome model $(Y|A, M, X_L)$, mediation model $(M|A, X_L)$ and treatment model $(A|X_L)$, thus penalizing a variable that is likely not a parent of the dependent variable.

In the sequel, we present the mglasso algorithm.

Algorithm The proposed ***mglasso***

1: Given original data $\mathbf{X} = (\mathbf{X}_L, A, M, Y) \in \mathbb{R}^{q=p+3}$;

2: Start with $\mathbf{W} = \mathbf{S} + \mathbf{P}_\rho$;

3: Solve the lasso problem (4.6) and save the value of $\hat{\boldsymbol{\beta}}$;

4: Update $\mathbf{w}_{12}$ and $\mathbf{w}_{21}$ using the relation $\mathbf{w}_{12} = \mathbf{W}_{11}\hat{\boldsymbol{\beta}}$;

5: Rearrange $\mathbf{W}$ so the next row and column are in position $i = 1, j = 2$;

6: Repeat steps 2-5 until convergence ;

7: Calculate the diagonal of $\boldsymbol{\Theta}$ using $\hat{\theta}_{22} = 1/(w_{22} - \mathbf{w}_{12}^T\hat{\boldsymbol{\beta}})$;

8: Use the most recent value of $\boldsymbol{\beta}$ to complete $\boldsymbol{\Theta}$ (using $\boldsymbol{\beta} = -\frac{\boldsymbol{\theta}_{12}}{\theta_{22}}$).

Tuning parameter selection

We follow Baldé et al. (2022) to set the possible values of the tuning parameter ρ (S_ρ):

$$S_\rho = \{0, 10^{-2}, 10^{-1.5}, 10^{-1}, 10^{-0.75}, 10^{-0.5}, 10^{-0.25}, 10^{0}, 10^{0.25}, 10^{0.5}, 10^{1}\}.$$

The set S_ρ was similar to a set of possible values of the tuning parameter proposed in Zou and Hastie (2005). As indicated before, we use the penalty $\|\boldsymbol{\Theta} * \mathbf{P}_\rho\|_1$ instead of $\rho\|\boldsymbol{\Theta}\|_1$, where

$$\boldsymbol{\Theta} = \begin{bmatrix} \Theta_{11} & \Theta_{12} & \dots & \Theta_{1p} & \Theta_{1A} & \Theta_{1M} & \Theta_{1Y} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \Theta_{p1} & \Theta_{p2} & \dots & \Theta_{pp} & \Theta_{pA} & \Theta_{pM} & \Theta_{pY} \\ \Theta_{A1} & \Theta_{A2} & \dots & \Theta_{Ap} & \Theta_{AA} & \Theta_{AM} & \Theta_{AY} \\ \Theta_{M1} & \Theta_{M2} & \dots & \Theta_{Mp} & \Theta_{MA} & \Theta_{MM} & \Theta_{MY} \\ \Theta_{Y1} & \Theta_{Y2} & \dots & \Theta_{Yp} & \Theta_{YA} & \Theta_{YM} & \Theta_{YY} \end{bmatrix}$$

and

$$\mathbf{P}_\rho = \begin{bmatrix} \rho & \rho^{-1} & \dots & \rho^{-1} & \rho|\alpha_1|^{-\gamma} & \rho|\eta_1|^{-\gamma} & \rho|\delta_1|^{-\gamma} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho^{-1} & \rho^{-1} & \dots & \rho & \rho|\alpha_p|^{-\gamma} & \rho|\eta_p|^{-\gamma} & \rho|\delta_p|^{-\gamma} \\ \rho|\alpha_1|^{-\gamma} & \rho|\alpha_2|^{-\gamma} & \dots & \rho|\alpha_p|^{-\gamma} & \rho & \rho|\eta_A|^{-\gamma} & \rho|\delta_A|^{-\gamma} \\ \rho|\eta_1|^{-\gamma} & \rho|\eta_2|^{-\gamma} & \dots & \rho|\eta_p|^{-\gamma} & \rho|\eta_A|^{-\gamma} & \rho & \rho|\delta_M|^{-\gamma} \\ \rho|\delta_1|^{-\gamma} & \rho|\delta_2|^{-\gamma} & \dots & \rho|\delta_p|^{-\gamma} & \rho|\delta_A|^{-\gamma} & \rho|\delta_M|^{-\gamma} & \rho \end{bmatrix},$$

where $\delta_L = (\delta_1, \dots, \delta_p)$ are the outcome regression coefficients for the covariates, $\eta_L = (\eta_1, \dots, \eta_p)$ are the mediator regression coefficients for the covariates and $\alpha_L = (\alpha_1, \dots, \alpha_p)$ are the treatment regression coefficients of the covariates; δ_A and δ_M are the outcome regression coefficients of A and M, respectively, and η_A the mediator regression coefficient of A. The tuning parameter ρ is estimated with a K-fold ($K = 5$) cross validation based on the BIC criterion using the `CVglasso` function (Friedman *et al.*, 2008; Galloway, 2018). Following the discussion in Shortreed and Ertefaie (2017), we use $\gamma = 2$.

4.3 Simulation

In this section, we describe the simulation study conceived for examining the usefulness and performance of the glasso and proposed mglasso approaches in simple mediation scenarios. We considered data generation scenarios inspired by Baldé and Lefebvre (2022+) and Ye et al. (2021). We present the simulation setup used for generating the data $(\mathbf{L}, A, M, Y)$, denoting the matrix containing all covariates, treatment (exposure), mediator and outcome, respectively. We generated the vector $L_i = (L_{i1}, L_{i2}, \dots, L_{ip})_{1 \leq i \leq n}$ from a multivariate standard Gaussian distribution with independant covariates ($L_i \sim N_p(0, \mathbf{I})$). The continuous treatment A_i was simulated as $A_i = \sum_{j=1}^{p} \alpha_j L_{ij} + \epsilon_i^A$, $\epsilon_i^A \sim N(0, 1)$. The continuous mediator M_i was generated as $M_i = \eta_A A_i + \sum_{j=1}^{p} \eta_j L_{ij} + \epsilon_i^M$, $\epsilon_i^M \sim N(0, 1)$ and the continuous outcome Y_i as $Y_i = \delta_A A_i + \delta_M M_i + \sum_{j=1}^{p} \delta_j L_j + \epsilon_i^Y$, $\epsilon_i^Y \sim N(0, 1)$. The true direct effect (DE) value was $\delta_A = 1.5$, the true indirect effect (IE) was $\delta_M \times \eta_A = 2 \times 1 = 2$ and the true total effect (TE) was DE + IE $= 1.5 + 2 = 3.5$. We considered two mediation graph

structures (Case 1 and Case 2), and for each of them, three scenarios that differed according to the values of the regression coefficients in the outcome model, while the regression coefficients in the treatment and mediator models were held fixed.

Case 1: Small number of covariates ($p = 3$)

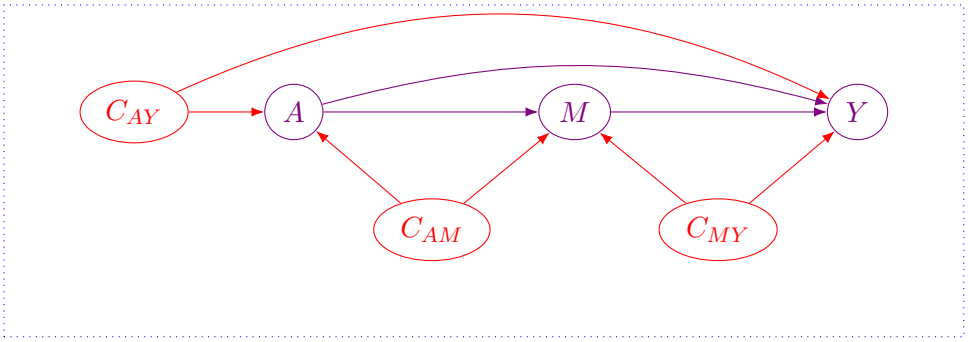


Figure 4.2 Hypothesized causal mediation graph with $p = 3$ covariates (confounders).

For the simplest graph structure with $p = 3$ covariates (Case 1, see Figure 4.2), the outcome regression coefficients $\delta_L \in \mathbb{R}^p$, the mediator regression coefficients $\eta_L \in \mathbb{R}^p$ and the treatment regression coefficients $\alpha_L \in \mathbb{R}^p$ were defined as follows:

1. Weakly related (δ_L): $\delta_L = (0.2, 0.2, 0)$, $\eta_L = (0, 1, 1)$, $\alpha_L = (1, 0, 1)$;
2. Moderately related (δ_L): $\delta_L = (0.6, 0.6, 0)$, $\eta_L = (0, 1, 1)$, $\alpha_L = (1, 0, 1)$;
3. Strongly related (δ_L): $\delta_L = (1, 1, 0)$, $\eta_L = (0, 1, 1)$, $\alpha_L = (1, 0, 1)$.

Case 2: Modest number of covariates ($p = 10$)

For the more complex graph structure with $p = 10$ covariates (Case 2, see Figure 4.1), the regression coefficients were defined as follows:

1. Weakly related (δ_L): $\delta_L = (0.2, 0.2, 0, 0.2, 0, 0, 0, \ldots, 0)$, $\eta_L = (0, 1, 1, 0, 0, 1, 0, \ldots, 0)$, $\alpha_L = (1, 0, 1, 0, 1, 0, 0, \ldots, 0)$;

2. Moderately related (δ_L): $\delta_L = (0.6, 0.6, 0, 0.6, 0, 0, 0, \ldots, 0)$, $\eta_L = (0, 1, 1, 0, 0, 1, 0, \ldots, 0)$, $\alpha_L = (1, 0, 1, 0, 1, 0, 0, \ldots, 0)$;

3. Strongly related (δ_L): $\delta_L = (1, 1, 0, 1, 0, 0, 0, \ldots, 0)$, $\eta_L = (0, 1, 1, 0, 0, 1, 0, \ldots, 0)$, $\alpha_L = (1, 0, 1, 0, 1, 0, 0, \ldots, 0)$.

In both cases and each scenario, the first covariate is a treatment-outcome confounder, the second is a mediator-outcome confounder, the third is a treatment-mediator confounder. In each scenario of Case 2, the fourth covariate is a pure predictor of outcome, the fifth covariate is a pure predictor of treatment, the sixth covariate is a pure predictor of mediator and the rest of covariates are spurious ($p = 10$; $p - 6 = 4$). We simulated 1000 data sets of sizes $n = 200, 500$ for each scenario.

Because treatment variables are often binary, we redid all these simulations using a binary treatment variable instead of a continuous one. The binary treatment A_i was simulated from a Bernoulli distribution with $logit\{P(A_i = 1|L)\} = \sum_{j=1}^{p} \alpha_j L_{ij}$.

Lastly, we note that in our implementation of mglasso and glasso in the simulations, we take $X_L \equiv L$. This allows to easily identify missing and spurious edges between covariates, and assess whether their types can be correctly determined from the algorithms.

4.4 Results

Figures 4.3-4.6 and Web Tables 4.1-4.12 present mglasso results associated with Scenarios 1-3 for both cases (Case 1: small number of covariates ($p = 3$) and Case 2: modest ($p = 10$) number of covariates), and both sample sizes investigated ($n = 200, 500$). These figures display on the left the true skeleton graphs corresponding to the mediation DAGs and, on the right, the estimated network graphs with the proposed mglasso (in these graphs, an edge is drawn if it is selected in $\geq 50\%$ of Monte-Carlo replications).

Figures 4.7-4.10 and Web Tables 4.13-4.24 present glasso results associated with Scenarios 1-3 for both cases (Case 1: small number of covariates ($p = 3$) and Case 2: modest ($p = 10$) number of covariates), and both sample sizes investigated ($n = 200, 500$). These figures display on the left the true moral graphs corresponding to the mediation DAGs (obtained using `DAGitty R` package) (Textor *et al.*, 2016) and, on the right,

the estimated network graphs with glasso (in these graphs, an edge is drawn if it is selected in $\geq 50\%$ of Monte-Carlo replications). As one can notice, the moral graphs are significantly more complex than the skeleton graphs even in these simple artificial scenarios.

All the results for the binary treatment simulations are presented in Web Appendix B (Web Figures 4.11-4.12). Due to space constraints, we omit the presentation of the tables containing the proportion of times each edge was selected over 1000 replications. Figure 4.3 shows that the graph estimated by mglasso was the same as the true skeleton graph when $n = 200$ and $p = 3$ (Case 1) in all scenarios. For Case 1, Figure 4.4 shows the exact same estimated graph for mglasso as in Figure 4.3 in all scenarios, when the sample size was increased to $n = 500$. For Case 2 ($p = 10$), Figure 4.5 shows that mglasso misses three edges of the true skeleton graph when $n = 200$. In Scenario 1 (when $n = 200$), mglasso missed the three edges related to the outcome ($Y \longleftrightarrow C_{AY}$, $Y \longleftrightarrow C_{MY}$ and $Y \longleftrightarrow P_Y$), which can be explained by the fact that the outcome regression coefficients δ_L are weak in this scenario. For that scenario, Web Table 4.7 shows that the edges $Y \longleftrightarrow C_{AY}$, $Y \longleftrightarrow C_{MY}$ and $Y \longleftrightarrow P_Y$ were selected about 0%, 39% and 0%, respectively. However, mglasso was able to capture the true skeleton graph when the sample size was increased to $n = 500$ in Scenarios 2-3 and missed only the edge $Y \longleftrightarrow P_Y$ in Scenario 1 (see Figure 4.6). Web Table 4.10 shows that the edge $Y \longleftrightarrow P_Y$ was selected about 0% of times in that scenario. Taking into account the central and known role of variables A, M and Y in mediation analyses, these results show that mglasso can facilitate the identification of covariate' types ($C_{AY}, C_{MY}, C_{AM}, P_Y, P_M, P_A$).

Figure 4.7 shows that the graph estimated by glasso is not the same as the true moral graph when $n = 200$ and $p = 3$ (Case 1) in all scenarios. While the moral graph features no relationship between the outcome and the confounder of the mediator-treatment relationship ($Y \longleftrightarrow C_{AM}$), glasso identifies this edge about 98% of times (see Web Tables 4.13-4.15). We observe the same phenomenon with a sample size of $n = 500$, as the edge $Y \longleftrightarrow C_{AM}$ was selected about 100% of times (see Web Tables 4.16-4.18). For Case 2 (when $p = 10$), differences between the moral graph and the graph estimated by

glasso were more important (see Figures 4.9 and 4.10). In this case, glasso introduced more and more spurious edges as the strength of the outcome regression coefficents and sample size increased.

When a binary treatment was considered in Case 1 ($p = 3$), the mglasso captured the true skeleton graph in all scenarios and for all sample sizes ($n = 200, 500$). This result mimics what was obtained with a continuous treatment. In Case 2 ($p = 10$) with $n = 200$, mglasso with a binary treatment added the incorrect edge $A \longleftrightarrow P_M$ to the graph selected by mglasso when a continuous treatment was used in Scenarios 1 and 3 with $n = 200$ (see Web Figure 4.11). When the sample size was increased to $n = 500$, mglasso with a continuous treatment captured the true skeleton graph in Scenarios 2-3. However, for this sample size, mglasso with a binary treatment selected the same graph as mglasso with a continuous treatment when the sample size was $n = 200$ (see Web Figure 4.12).

4.5 Discussion

In this article, we proposed a graphical lasso algorithm (mglasso) to learn causal mediation graphs from observed multivariate normal data. The glasso is a very popular and fast algorithm used to estimate the precision matrix which encodes conditional dependencies between Gaussian variables. However, as seen in this work, even very simple causal mediation DAGs yield networks with complex dependencies which can be hard to interpret in practice. The complexity of these networks is notably the result of dependencies induced by conditioning on common effects. In other words, it is expected that several true edges between variables exist because conditioning is done on colliders, which inevitably obscurs interpretation. Our approach uses componentwise adaptive weights for the treatment, mediator and outcome variables, which are defined using the treatment model, mediator model and outcome model regression coefficients to penalize weak associations. Note that each of these models only conditions on variables presumably antecedent to the dependent variable in each case, that is, it exploits posited knowledge on temporal ordering of variables in mediation (covariates, treatment, mediator, outcome). Through this, our proposed glasso approach is thus able to reveal direct associations between variables in resulting network and encourage "collider-free sparsity". The simulation results showed that the mglasso algorithm is able to identify the true skeleton graph of a DAG, which is arguably more useful for interpretation and to identify covariates types than the true moral graph as shown in a dependencies network. In our simulation study, we performed all simulations with a small and modest number of covariates (low-dimensional setting). We will investigate high-dimensional scenarios (number of covariates increasing with the sample size) in a future work. We also leave the investigation of the asymptotic behavior of the proposed mglasso for future study. In addition, a simple graphical interface is being produced based on `R` Shiny application. This could be very useful in practice and serve as a tool for visualizing probable direct and mediated relationships between variables.

Figure 4.3 Causal mediation network estimation using the proposed mglasso under Scenarios 1, 2 and 3 (by row) in the low-dimensional setting with $p = 3$ (Case 1) and $n = 200$ (based on 1000 replications).

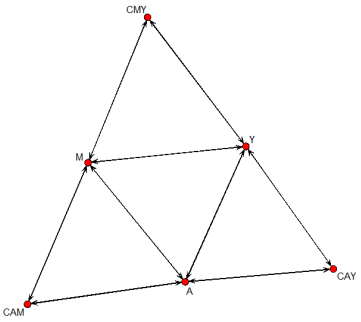
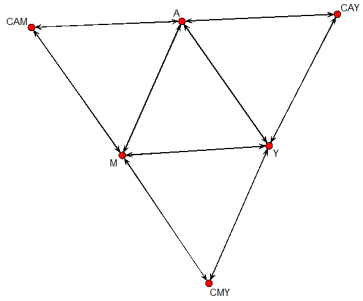
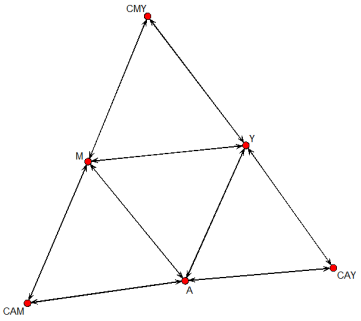
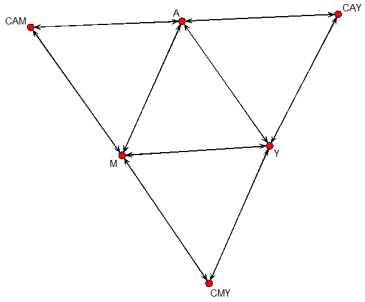
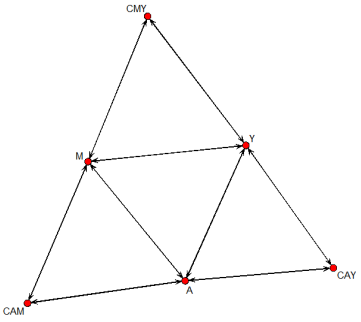
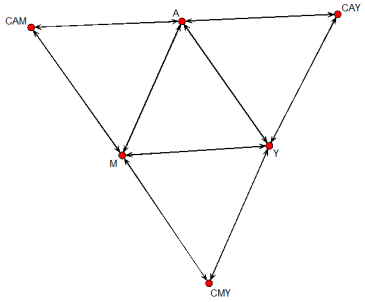
	skeleton graph	**mglasso graph**
1		
2		
3		

Figure 4.4 Causal mediation network estimation using the proposed mglasso under Scenarios 1, 2 and 3 (by row) in the low-dimensional setting with $p = 3$ (Case 1) and $n = 500$ (based on 1000 replications).

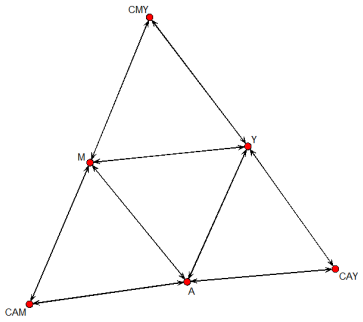
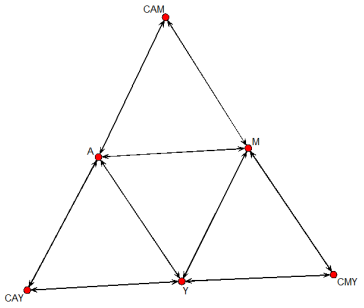
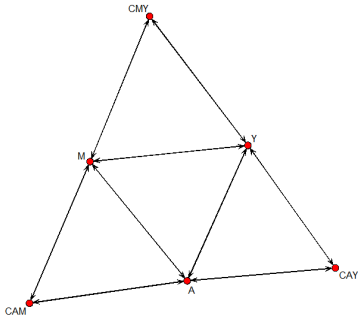
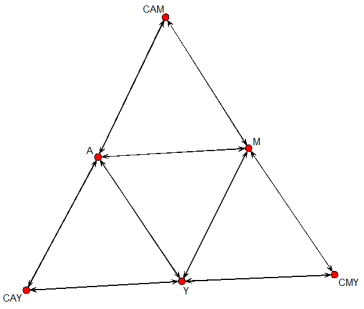
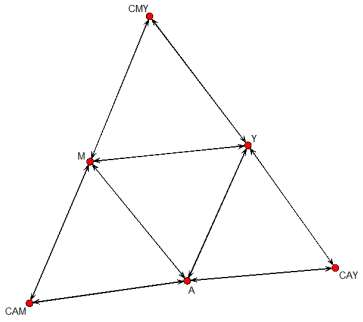
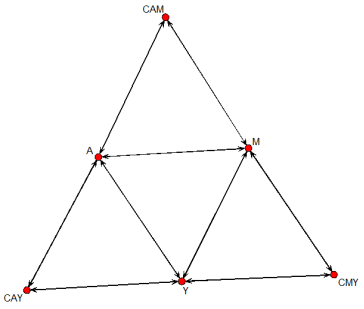
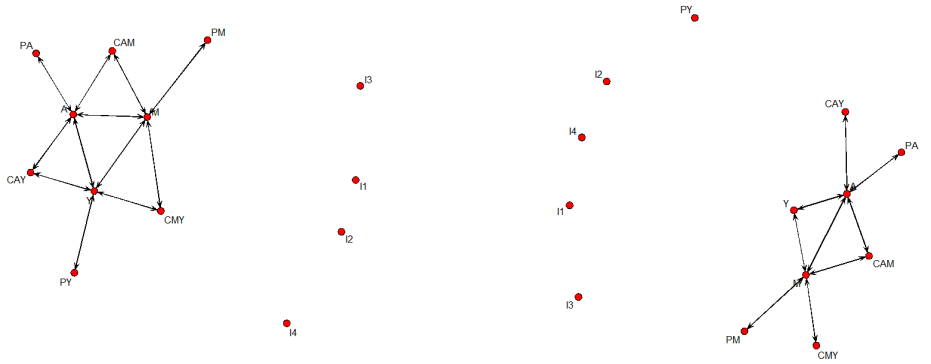
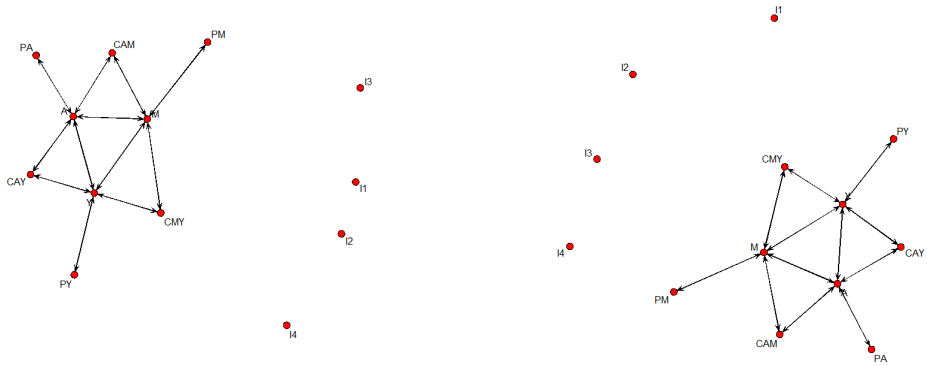
	skeleton graph	**mglasso graph**
1		
2		
3		

Figure 4.5 Causal mediation network estimation using the proposed mglasso under Scenarios 1, 2 and 3 (by row) in the low-dimensional setting with $p = 10$ (Case 2) and $n = 200$ (based on 1000 replications).

skeleton graph **mglasso graph**



1

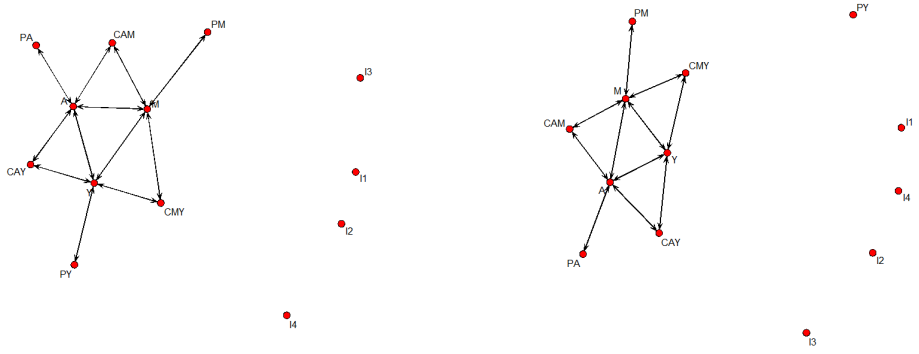


2

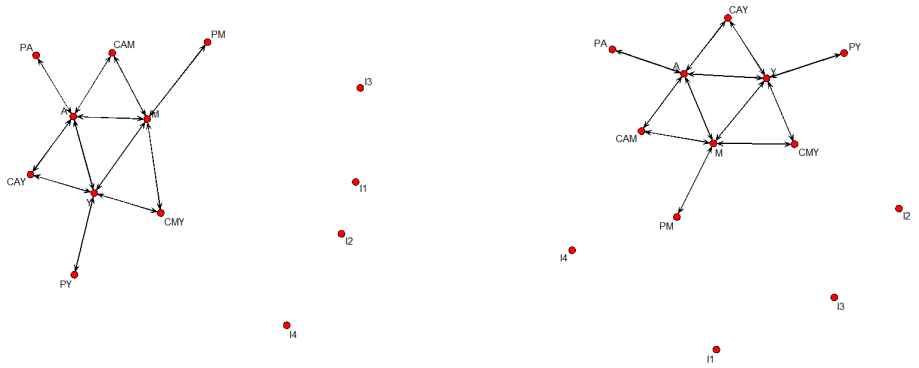
3

Figure 4.6 Causal mediation network estimation using the proposed mglasso under Scenarios 1, 2 and 3 (by row) in the low-dimensional setting with $p = 10$ (Case 2) and $n = 500$ (based on 1000 replications).

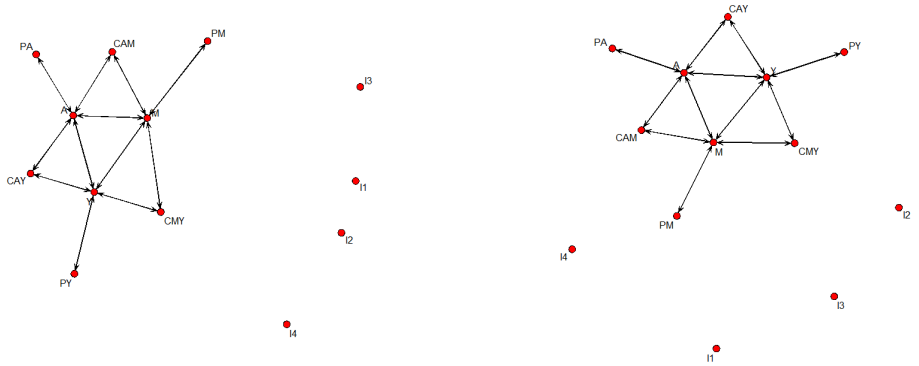
skeleton graph **mglasso graph**



1



2



3

Figure 4.7 Causal mediation network estimation using glasso under Scenarios 1, 2 and 3 (by row) in the low-dimensional setting with $p = 3$ (Case 1) and $n = 200$ (based on 1000 replications).

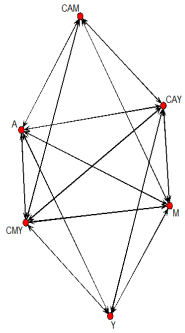
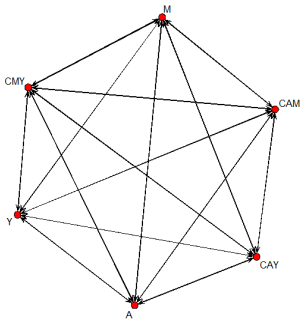
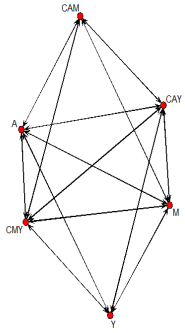
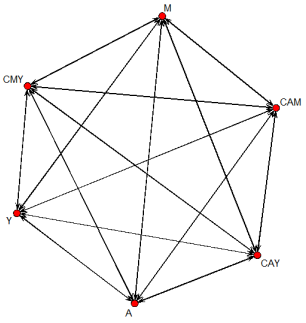
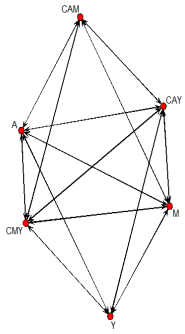
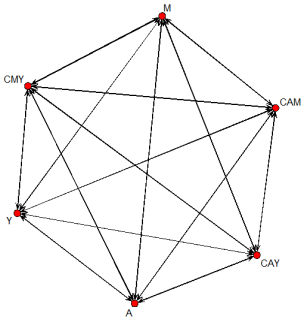
	moral graph	**glasso graph**
1		
2		
3		

Figure 4.8 Causal mediation network estimation using glasso under Scenarios 1, 2 and 3 (by row) in the low-dimensional setting with $p = 3$ (Case 1) and $n = 500$ (based on 1000 replications).

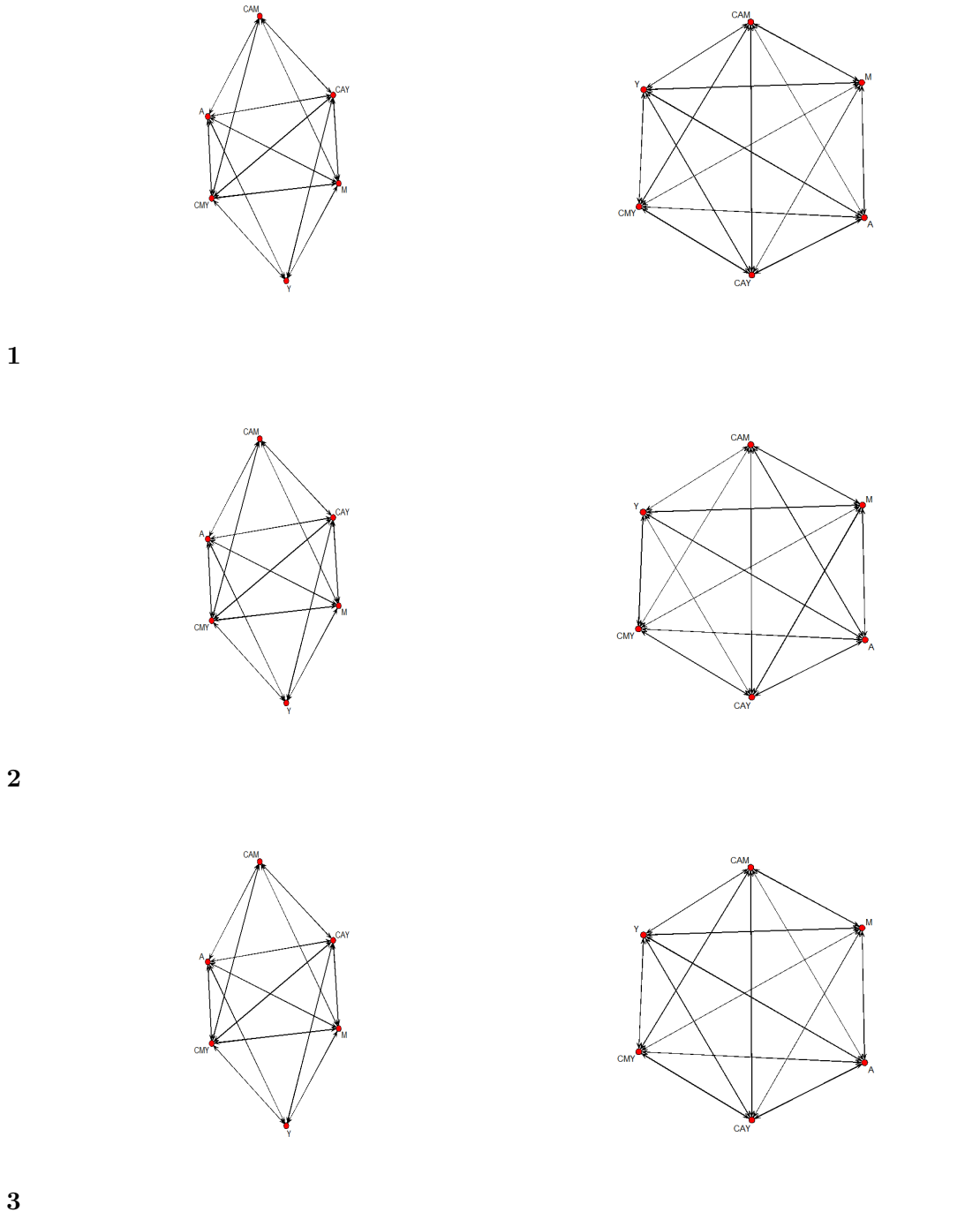


Figure 4.9 Causal mediation network estimation using glasso under Scenarios 1, 2 and 3 (by row) in the low-dimensional setting with $p = 10$ (Case 2) and $n = 200$ (based on 1000 replications).

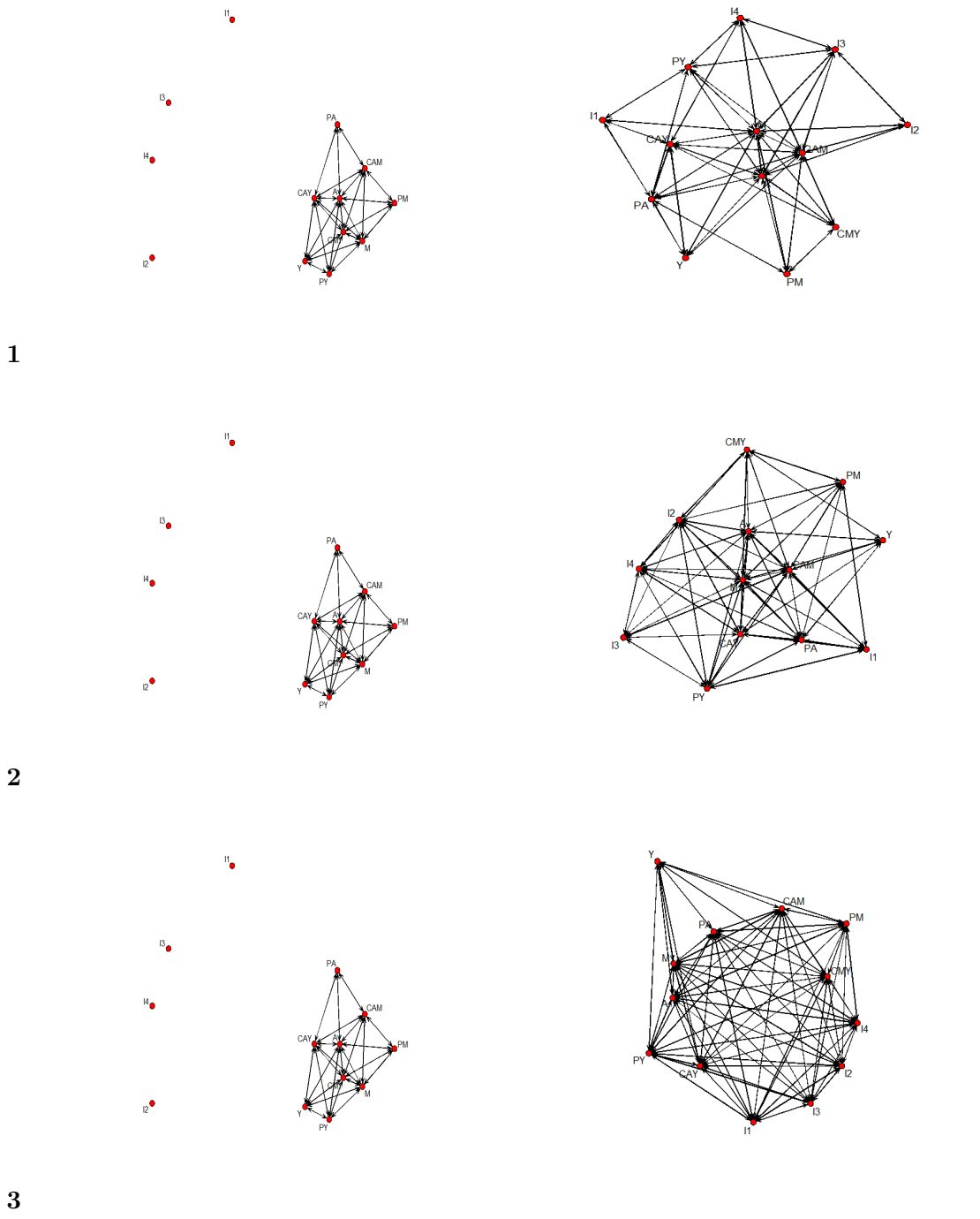
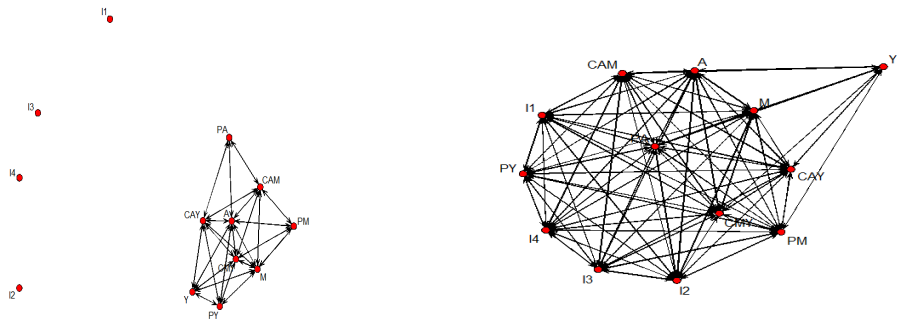
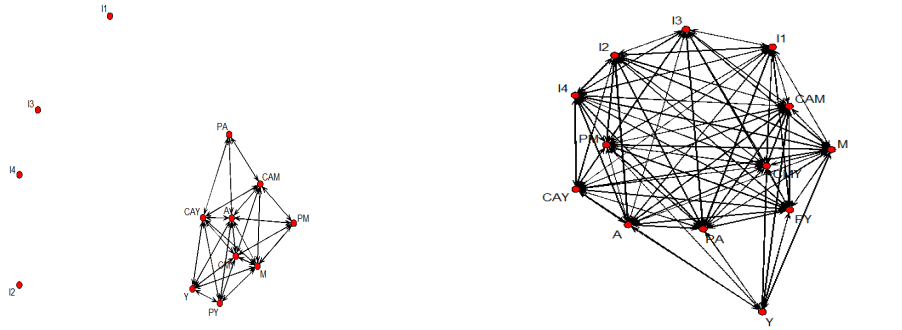


Figure 4.10 Causal mediation network estimation using glasso under Scenarios 1, 2 and 3 (by row) in the low-dimensional setting with $p = 10$ (Case 2) and $n = 500$ (based on 1000 replications).

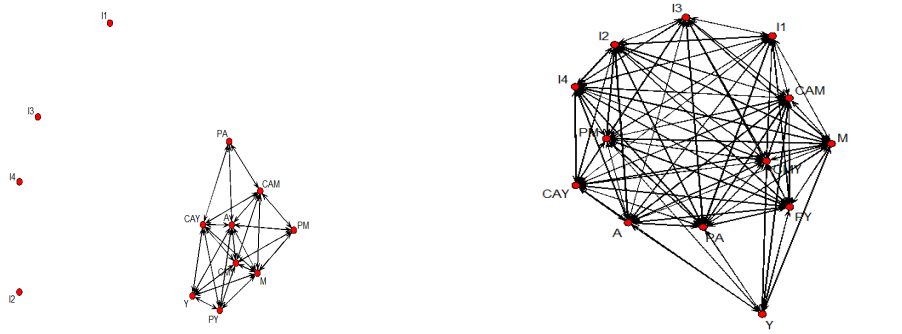
moral graph **glasso graph**



1



2



3

4.6 Web Appendix

4.6.1 Web Appendix A. Mean adjacency matrices

Web Tables 4.1-4.12 present the proportion of times each edge was selected over 1000 replications by the proposed mglasso. Web Tables 4.13-4.24 present the proportion of times each edge was selected over 1000 replications by the glasso. All these tables correspond to results obtained from simulations with a continuous treatment variable.

Tableau 4.1 Probability of edge being selected based on 1000 replications using the proposed mglasso under Scenario 1 with sample size $n = 200$ and $p = 3$ covariates.

	C_{AY}	C_{MY}	C_{AM}	A	M	Y
C_{AY}	0.00	0.00	0.00	1.00	0.00	0.71
C_{MY}	0.00	0.00	0.00	0.00	0.98	0.89
C_{AM}	0.00	0.00	0.00	0.96	0.99	0.00
A	1.00	0.00	0.96	0.00	0.99	1.00
M	0.00	0.98	0.99	0.99	0.00	1.00
Y	0.71	0.89	0.00	1.00	1.00	0.00

Tableau 4.2 Probability of edge being selected based on 1000 replications using the proposed mglasso under Scenario 2 with sample size $n = 200$ and $p = 3$ covariates.

	C_{AY}	C_{MY}	C_{AM}	A	M	Y
C_{AY}	0.00	0.00	0.00	0.99	0.00	0.99
C_{MY}	0.00	0.00	0.00	0.00	0.98	0.99
C_{AM}	0.00	0.00	0.00	0.97	0.99	0.00
A	0.99	0.00	0.97	0.00	0.99	1.00
M	0.00	0.98	0.99	0.99	0.00	1.00
Y	0.99	0.99	0.00	1.00	1.00	0.00

4.6.2 Web Appendix B. Additional simulation results with binary treatment

Web Figures 4.11-4.12 present results for both mglasso and glasso with binary treatment under Scenarios 1-3 when $p = 10$.

Tableau 4.3 Probability of edge being selected based on 1000 replications using the proposed mglasso under Scenario 3 with sample size $n = 200$ and $p = 3$ covariates.

	C_{AY}	C_{MY}	C_{AM}	A	M	Y
C_{AY}	0.00	0.00	0.00	0.99	0.00	1.00
C_{MY}	0.00	0.00	0.00	0.00	0.99	1.00
C_{AM}	0.00	0.00	0.00	0.97	0.99	0.00
A	0.99	0.00	0.97	0.00	1.00	1.00
M	0.00	0.99	0.99	1.00	0.00	1.00
Y	1.00	1.00	0.00	1.00	1.00	0.00

Tableau 4.4 Probability of edge being selected based on 1000 replications using the proposed mglasso under Scenario 1 with sample size $n = 500$ and $p = 3$ covariates.

	C_{AY}	C_{MY}	C_{AM}	A	M	Y
C_{AY}	0.00	0.00	0.00	1.00	0.00	0.88
C_{MY}	0.00	0.00	0.00	0.00	1.00	0.88
C_{AM}	0.00	0.00	0.00	1.00	1.00	0.00
A	1.00	0.00	1.00	0.00	1.00	1.00
M	0.00	1.00	1.00	1.00	0.00	1.00
Y	0.88	0.88	0.00	1.00	1.00	0.00

Tableau 4.5 Probability of edge being selected based on 1000 replications using the proposed mglasso under Scenario 2 with sample size $n = 500$ and $p = 3$ covariates.

	C_{AY}	C_{MY}	C_{AM}	A	M	Y
C_{AY}	0.00	0.00	0.00	1.00	0.00	1.00
C_{MY}	0.00	0.00	0.00	0.00	1.00	1.00
C_{AM}	0.00	0.00	0.00	1.00	1.00	0.00
A	1.00	0.00	1.00	0.00	1.00	1.00
M	0.00	1.00	1.00	1.00	0.00	1.00
Y	1.00	1.00	0.00	1.00	1.00	0.00

Tableau 4.6 Probability of edge being selected based on 1000 replications using the proposed mglasso under Scenario 3 with sample size $n = 500$ and $p = 3$ covariates.

	C_{AY}	C_{MY}	C_{AM}	A	M	Y
C_{AY}	0.00	0.00	0.00	1.00	0.00	1.00
C_{MY}	0.00	0.00	0.00	0.00	1.00	1.00
C_{AM}	0.00	0.00	0.00	1.00	1.00	0.00
A	1.00	0.00	1.00	0.00	1.00	1.00
M	0.00	1.00	1.00	1.00	0.00	1.00
Y	1.00	1.00	0.00	1.00	1.00	0.00

Tableau 4.7 Probability of edge being selected based on 1000 replications using the proposed mglasso under Scenario 1 with sample size $n = 200$ and $p = 10$ covariates.

	C_{AY}	C_{MY}	C_{AM}	P_Y	P_A	P_M	I_1	I_2	I_3	I_4	A	M	Y
C_{AY}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.06	0.00
C_{MY}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.93	0.39
C_{AM}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.82	1.00	0.00
P_Y	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
P_A	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.37	0.00
P_M	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	1.00	0.00
I_1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
I_2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
I_3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02
I_4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
A	1.00	0.00	0.82	0.00	1.00	0.01	0.00	0.00	0.00	0.00	0.00	0.82	1.00
M	0.06	0.93	1.00	0.00	0.37	1.00	0.00	0.01	0.00	0.00	0.82	0.00	1.00
Y	0.00	0.39	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	1.00	1.00	0.00

Tableau 4.8 Probability of edge being selected based on 1000 replications using the proposed mglasso under Scenario 2 with sample size $n = 200$ and $p = 10$ covariates.

	C_{AY}	C_{MY}	C_{AM}	P_Y	P_A	P_M	I_1	I_2	I_3	I_4	A	M	Y
C_{AY}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.93	0.00	0.76
C_{MY}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.70	0.92
C_{AM}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.83	1.00	0.00
P_Y	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.87
P_A	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.40	0.00
P_M	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
I_1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
I_2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
I_3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02
I_4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
A	0.93	0.00	0.83	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.78	1.00
M	0.00	0.70	1.00	0.00	0.40	1.00	0.00	0.01	0.00	0.00	0.78	0.00	1.00
Y	0.76	0.92	0.00	0.87	0.00	0.00	0.00	0.00	0.02	0.00	1.00	1.00	0.00

Tableau 4.9 Probability of edge being selected based on 1000 replications using the proposed mglasso under Scenario 3 with sample size $n = 200$ and $p = 10$ covariates.

	C_{AY}	C_{MY}	C_{AM}	P_Y	P_A	P_M	I_1	I_2	I_3	I_4	A	M	Y
C_{AY}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.92	0.00	1.00
C_{MY}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.80	1.00
C_{AM}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.87	1.00	0.00
P_Y	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
P_A	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.45	0.00
P_M	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	1.00	0.00
I_1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
I_2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00
I_3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02
I_4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
A	0.92	0.00	0.87	0.00	1.00	0.01	0.00	0.00	0.00	0.00	0.00	0.86	1.00
M	0.00	0.80	1.00	0.00	0.45	1.00	0.00	0.02	0.00	0.00	0.86	0.00	1.00
Y	1.00	1.00	0.00	1.00	0.00	0.00	0.00	0.00	0.02	0.00	1.00	1.00	0.00

Tableau 4.10 Probability of edge being selected based on 1000 replications using the proposed mglasso under Scenario 1 with sample size $n = 500$ and $p = 10$ covariates.

	C_{AY}	C_{MY}	C_{AM}	P_Y	P_A	P_M	I_1	I_2	I_3	I_4	A	M	Y
C_{AY}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.93
C_{MY}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.92
C_{AM}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	1.00	0.00
P_Y	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
P_A	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
P_M	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
I_1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
I_2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
I_3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
I_4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
A	1.00	0.00	0.99	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00
M	0.00	1.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00
Y	0.93	0.92	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	0.00

Tableau 4.11 Probability of edge being selected based on 1000 replications using the proposed mglasso under Scenario 2 with sample size $n = 500$ and $p = 10$ covariates.

	C_{AY}	C_{MY}	C_{AM}	P_Y	P_A	P_M	I_1	I_2	I_3	I_4	A	M	Y
C_{AY}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00
C_{MY}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00
C_{AM}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	1.00	0.00
P_Y	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
P_A	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
P_M	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
I_1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
I_2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
I_3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
I_4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
A	1.00	0.00	0.99	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00
M	0.00	1.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00
Y	1.00	1.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	0.00

Tableau 4.12 Probability of edge being selected based on 1000 replications using the proposed mglasso under Scenario 3 with sample size $n = 500$ and $p = 10$ covariates.

	C_{AY}	C_{MY}	C_{AM}	P_Y	P_A	P_M	I_1	I_2	I_3	I_4	A	M	Y
C_{AY}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00
C_{MY}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00
C_{AM}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	0.00
P_Y	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
P_A	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
P_M	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
I_1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
I_2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
I_3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
I_4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
A	1.00	0.00	1.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00
M	0.00	1.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00
Y	1.00	1.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	0.00

Tableau 4.13 Probability of edge being selected based on 1000 replications using the glasso under Scenario 1 with sample size $n = 200$ and $p = 3$ covariates.

	C_{AY}	C_{MY}	C_{AM}	A	M	Y
C_{AY}	0.00	0.84	0.90	1.00	0.82	0.98
C_{MY}	0.84	0.00	0.99	0.99	0.99	0.98
C_{AM}	0.90	0.99	0.00	0.98	0.99	0.98
A	1.00	0.99	0.98	0.00	1.00	1.00
M	0.82	0.99	0.99	1.00	0.00	1.00
Y	0.98	0.98	0.98	1.00	1.00	0.00

Tableau 4.14 Probability of edge being selected based on 1000 replications using the glasso under Scenario 2 with sample size $n = 200$ and $p = 3$ covariates.

	C_{AY}	C_{MY}	C_{AM}	A	M	Y
C_{AY}	0.00	0.84	0.91	0.99	0.83	0.99
C_{MY}	0.84	0.00	1.00	0.99	0.98	0.99
C_{AM}	0.91	1.00	0.00	0.99	0.99	0.98
A	0.99	0.99	0.99	0.00	1.00	1.00
M	0.83	0.98	0.99	1.00	0.00	1.00
Y	0.99	0.99	0.98	1.00	1.00	0.00

Tableau 4.15 Probability of edge being selected based on 1000 replications using the glasso under Scenario 3 with sample size $n = 200$ and $p = 3$ covariates.

	C_{AY}	C_{MY}	C_{AM}	A	M	Y
C_{AY}	0.00	0.89	0.93	0.99	0.81	1.00
C_{MY}	0.89	0.00	1.00	1.00	0.99	1.00
C_{AM}	0.93	1.00	0.00	0.99	1.00	0.98
A	0.99	1.00	0.99	0.00	1.00	1.00
M	0.81	0.99	1.00	1.00	0.00	1.00
Y	1.00	1.00	0.98	1.00	1.00	0.00

Tableau 4.16 Probability of edge being selected based on 1000 replications using the glasso under Scenario 1 with sample size $n = 500$ and $p = 3$ covariates.

	C_{AY}	C_{MY}	C_{AM}	A	M	Y
C_{AY}	0.00	1.00	0.80	1.00	0.77	1.00
C_{MY}	1.00	0.00	0.99	1.00	1.00	1.00
C_{AM}	0.80	0.99	0.00	1.00	1.00	1.00
A	1.00	1.00	1.00	0.00	1.00	1.00
M	0.77	1.00	1.00	1.00	0.00	1.00
Y	1.00	1.00	1.00	1.00	1.00	0.00

Tableau 4.17 Probability of edge being selected based on 1000 replications using the glasso under Scenario 2 with sample size $n = 500$ and $p = 3$ covariates.

	C_{AY}	C_{MY}	C_{AM}	A	M	Y
C_{AY}	0.00	1.00	0.82	1.00	0.70	1.00
C_{MY}	1.00	0.00	0.99	1.00	1.00	1.00
C_{AM}	0.82	0.99	0.00	1.00	1.00	1.00
A	1.00	1.00	1.00	0.00	1.00	1.00
M	0.70	1.00	1.00	1.00	0.00	1.00
Y	1.00	1.00	1.00	1.00	1.00	0.00

Tableau 4.18 Probability of edge being selected based on 1000 replications using the glasso under Scenario 3 with sample size $n = 500$ and $p = 3$ covariates.

	C_{AY}	C_{MY}	C_{AM}	A	M	Y
C_{AY}	0.00	1.00	0.82	1.00	0.77	1.00
C_{MY}	1.00	0.00	0.99	1.00	1.00	1.00
C_{AM}	0.82	0.99	0.00	1.00	1.00	1.00
A	1.00	1.00	1.00	0.00	1.00	1.00
M	0.77	1.00	1.00	1.00	0.00	1.00
Y	1.00	1.00	1.00	1.00	1.00	0.00

Tableau 4.19 Probability of edge being selected based on 1000 replications using the glasso under Scenario 1 with sample size $n = 200$ and $p = 10$ covariates.

	C_{AY}	C_{MY}	C_{AM}	P_Y	P_A	P_M	I_1	I_2	I_3	I_4	A	M	Y
C_{AY}	0.00	0.63	1.00	0.45	0.99	0.44	0.51	0.49	0.47	0.64	1.00	0.93	0.67
C_{MY}	0.63	0.00	1.00	0.45	0.46	0.99	0.44	0.50	0.43	0.45	1.00	0.99	0.47
C_{AM}	1.00	1.00	0.00	0.56	1.00	0.99	0.47	0.57	0.54	0.54	0.50	1.00	0.76
P_Y	0.45	0.45	0.56	0.00	0.51	0.42	0.75	0.49	0.51	0.57	0.55	0.74	0.09
P_A	0.99	0.46	1.00	0.51	0.00	0.65	0.60	0.48	0.46	0.47	1.00	0.91	0.71
P_M	0.44	0.99	0.99	0.42	0.65	0.00	0.50	0.50	0.42	0.44	1.00	1.00	0.45
I_1	0.51	0.44	0.47	0.75	0.60	0.50	0.00	0.44	0.41	0.41	0.53	0.76	0.07
I_2	0.49	0.50	0.57	0.49	0.48	0.50	0.44	0.00	0.51	0.48	0.76	0.59	0.10
I_3	0.47	0.43	0.54	0.51	0.46	0.42	0.41	0.51	0.00	0.55	0.53	0.89	0.10
I_4	0.64	0.45	0.54	0.57	0.47	0.44	0.41	0.48	0.55	0.00	0.49	0.56	0.11
A	1.00	1.00	0.50	0.55	1.00	1.00	0.53	0.76	0.53	0.49	0.00	0.96	1.00
M	0.93	0.99	1.00	0.74	0.91	1.00	0.76	0.59	0.89	0.56	0.96	0.00	1.00
Y	0.67	0.47	0.76	0.09	0.71	0.45	0.07	0.10	0.10	0.11	1.00	1.00	0.00

Tableau 4.20 Probability of edge being selected based on 1000 replications using the glasso under Scenario 2 with sample size $n = 200$ and $p = 10$ covariates.

	C_{AY}	C_{MY}	C_{AM}	P_Y	P_A	P_M	I_1	I_2	I_3	I_4	A	M	Y
C_{AY}	0.00	0.52	1.00	0.52	0.99	0.49	0.54	0.53	0.54	0.66	0.95	0.96	0.83
C_{MY}	0.52	0.00	1.00	0.50	0.48	0.99	0.47	0.55	0.48	0.51	1.00	0.63	0.66
C_{AM}	1.00	1.00	0.00	0.67	1.00	0.99	0.52	0.58	0.56	0.56	0.55	1.00	0.67
P_Y	0.52	0.50	0.67	0.00	0.51	0.46	0.77	0.54	0.53	0.60	0.95	1.00	0.36
P_A	0.99	0.48	1.00	0.51	0.00	0.66	0.61	0.51	0.50	0.50	1.00	0.84	0.67
P_M	0.49	0.99	0.99	0.46	0.66	0.00	0.54	0.53	0.46	0.48	1.00	1.00	0.50
I_1	0.54	0.47	0.52	0.77	0.61	0.54	0.00	0.50	0.46	0.46	0.57	0.71	0.08
I_2	0.53	0.55	0.58	0.54	0.51	0.53	0.50	0.00	0.54	0.51	0.77	0.61	0.12
I_3	0.54	0.48	0.56	0.53	0.50	0.46	0.46	0.54	0.00	0.59	0.56	0.90	0.09
I_4	0.66	0.51	0.56	0.60	0.50	0.48	0.46	0.51	0.59	0.00	0.52	0.56	0.15
A	0.95	1.00	0.55	0.95	1.00	1.00	0.57	0.77	0.56	0.52	0.00	0.96	1.00
M	0.96	0.63	1.00	1.00	0.84	1.00	0.71	0.61	0.90	0.56	0.96	0.00	1.00
Y	0.83	0.66	0.67	0.36	0.67	0.50	0.08	0.12	0.09	0.15	1.00	1.00	0.00

Tableau 4.21 Probability of edge being selected based on 1000 replications using the glasso under Scenario 3 with sample size $n = 200$ and $p = 10$ covariates.

	C_{AY}	C_{MY}	C_{AM}	P_Y	P_A	P_M	I_1	I_2	I_3	I_4	A	M	Y
C_{AY}	0.00	0.60	1.00	0.78	1.00	0.55	0.59	0.58	0.62	0.72	0.85	0.91	0.98
C_{MY}	0.60	0.00	1.00	0.74	0.55	0.99	0.51	0.60	0.54	0.56	1.00	0.64	0.99
C_{AM}	1.00	1.00	0.00	0.78	1.00	0.99	0.57	0.62	0.59	0.61	0.67	1.00	0.62
P_Y	0.78	0.74	0.78	0.00	0.56	0.50	0.84	0.59	0.57	0.68	0.97	1.00	0.92
P_A	1.00	0.55	1.00	0.56	0.00	0.73	0.68	0.56	0.54	0.54	1.00	0.80	0.65
P_M	0.55	0.99	0.99	0.50	0.73	0.00	0.60	0.58	0.51	0.53	1.00	1.00	0.53
I_1	0.59	0.51	0.57	0.84	0.68	0.60	0.00	0.54	0.50	0.51	0.64	0.67	0.12
I_2	0.58	0.60	0.62	0.59	0.56	0.58	0.54	0.00	0.61	0.56	0.79	0.65	0.12
I_3	0.62	0.54	0.59	0.57	0.54	0.51	0.50	0.61	0.00	0.65	0.60	0.91	0.10
I_4	0.72	0.56	0.61	0.68	0.54	0.53	0.51	0.56	0.65	0.00	0.55	0.55	0.22
A	0.85	1.00	0.67	0.97	1.00	1.00	0.64	0.79	0.60	0.55	0.00	0.85	1.00
M	0.90	0.64	1.00	1.00	0.80	1.00	0.67	0.65	0.91	0.55	0.85	0.00	1.00
Y	0.98	0.99	0.62	0.92	0.65	0.53	0.12	0.12	0.10	0.22	1.00	1.00	0.00

Tableau 4.22 Probability of edge being selected based on 1000 replications using the glasso under Scenario 1 with sample size $n = 500$ and $p = 10$ covariates.

	C_{AY}	C_{MY}	C_{AM}	P_Y	P_A	P_M	I_1	I_2	I_3	I_4	A	M	Y
C_{AY}	0.00	0.94	1.00	0.90	1.00	0.94	0.82	0.79	0.94	0.86	1.00	0.58	1.00
C_{MY}	0.94	0.00	1.00	0.82	0.98	1.00	0.79	0.86	0.86	0.81	1.00	1.00	0.98
C_{AM}	1.00	1.00	0.00	0.87	1.00	0.99	0.94	0.86	0.83	0.81	0.97	1.00	1.00
P_Y	0.90	0.82	0.87	0.00	0.80	0.82	0.82	0.81	0.78	0.84	0.87	0.88	0.34
P_A	1.00	0.98	1.00	0.80	0.00	0.99	0.90	0.80	0.92	0.78	1.00	0.98	1.00
P_M	0.94	1.00	0.99	0.82	0.99	0.00	0.87	0.99	0.78	0.83	1.00	1.00	0.98
I_1	0.82	0.79	0.94	0.82	0.90	0.87	0.00	0.84	0.79	0.96	0.79	0.91	0.10
I_2	0.79	0.86	0.86	0.81	0.80	0.99	0.84	0.00	0.80	0.81	0.80	0.79	0.13
I_3	0.94	0.86	0.83	0.78	0.92	0.78	0.79	0.80	0.00	0.85	0.83	0.85	0.15
I_4	0.86	0.81	0.81	0.84	0.78	0.83	0.96	0.81	0.85	0.00	0.78	0.85	0.32
A	1.00	1.00	0.97	0.87	1.00	1.00	0.79	0.80	0.83	0.78	0.00	0.99	1.00
M	0.58	1.00	1.00	0.88	0.98	1.00	0.91	0.79	0.85	0.85	0.99	0.00	1.00
Y	1.00	0.98	1.00	0.34	1.00	0.98	0.10	0.13	0.15	0.32	1.00	1.00	0.00

Tableau 4.23 Probability of edge being selected based on 1000 replications using the glasso under Scenario 2 with sample size $n = 500$ and $p = 10$ covariates.

	C_{AY}	C_{MY}	C_{AM}	P_Y	P_A	P_M	I_1	I_2	I_3	I_4	A	M	Y
C_{AY}	0.00	0.95	1.00	0.96	1.00	0.96	0.82	0.80	0.94	0.89	1.00	0.65	1.00
C_{MY}	0.95	0.00	1.00	0.82	0.99	1.00	0.82	0.88	0.87	0.83	1.00	0.99	0.99
C_{AM}	1.00	1.00	0.00	0.88	1.00	0.99	0.94	0.87	0.83	0.81	0.98	1.00	0.99
P_Y	0.96	0.82	0.88	0.00	0.80	0.84	0.83	0.81	0.78	0.84	0.98	0.99	0.93
P_A	1.00	0.99	1.00	0.80	0.00	0.99	0.91	0.82	0.93	0.79	1.00	0.97	0.99
P_M	0.96	1.00	0.99	0.84	0.99	0.00	0.88	0.99	0.80	0.84	1.00	1.00	0.96
I_1	0.82	0.81	0.94	0.83	0.91	0.88	0.00	0.84	0.80	0.96	0.80	0.89	0.09
I_2	0.80	0.88	0.87	0.81	0.82	0.99	0.84	0.00	0.81	0.82	0.80	0.82	0.11
I_3	0.94	0.87	0.83	0.78	0.93	0.80	0.80	0.81	0.00	0.86	0.83	0.85	0.15
I_4	0.89	0.83	0.81	0.84	0.79	0.84	0.96	0.82	0.86	0.00	0.78	0.85	0.33
A	1.00	1.00	0.98	0.98	1.00	1.00	0.80	0.80	0.83	0.78	0.00	0.99	1.00
M	0.65	0.99	1.00	0.99	0.97	1.00	0.89	0.82	0.85	0.85	0.99	0.00	1.00
Y	1.00	0.99	0.99	0.93	0.99	0.97	0.09	0.11	0.15	0.33	1.00	1.00	0.00

Tableau 4.24 Probability of edge being selected based on 1000 replications using the glasso under Scenario 3 with sample size $n = 500$ and $p = 10$ covariates.

	C_{AY}	C_{MY}	C_{AM}	P_Y	P_A	P_M	I_1	I_2	I_3	I_4	A	M	Y
C_{AY}	0.00	0.97	1.00	0.98	1.00	0.96	0.82	0.82	0.94	0.89	1.00	0.86	1.00
C_{MY}	0.97	0.00	1.00	0.80	0.99	1.00	0.80	0.87	0.88	0.81	1.00	0.99	1.00
C_{AM}	1.00	1.00	0.00	0.92	1.00	0.99	0.94	0.87	0.83	0.82	0.99	1.00	0.99
P_Y	0.98	0.80	0.92	0.00	0.81	0.84	0.83	0.80	0.79	0.85	1.00	1.00	1.00
P_A	1.00	0.99	1.00	0.81	0.00	0.99	0.91	0.81	0.93	0.78	1.00	0.96	0.99
P_M	0.96	1.00	0.99	0.84	0.99	0.00	0.88	0.99	0.80	0.85	1.00	1.00	0.95
I_1	0.82	0.80	0.94	0.83	0.91	0.88	0.00	0.84	0.80	0.96	0.81	0.87	0.09
I_2	0.82	0.87	0.87	0.80	0.81	0.99	0.84	0.00	0.81	0.83	0.80	0.83	0.10
I_3	0.94	0.88	0.83	0.79	0.93	0.80	0.80	0.81	0.00	0.86	0.84	0.85	0.15
I_4	0.89	0.81	0.82	0.85	0.78	0.84	0.96	0.83	0.86	0.00	0.78	0.84	0.32
A	1.00	1.00	0.99	1.00	1.00	1.00	0.81	0.80	0.84	0.77	0.00	0.99	1.00
M	0.86	0.99	1.00	1.00	0.96	1.00	0.87	0.83	0.85	0.84	0.99	0.00	1.00
Y	1.00	1.00	0.99	1.00	0.99	0.95	0.09	0.10	0.15	0.32	1.00	1.00	0.00

Figure 4.11 Causal mediation network estimation using both mglasso and glasso for binary treatment under Scenarios 1, 2 and 3 (by row) in the low-dimensional setting with $p = 10$ (Case 2) and $n = 200$ (based on 1000 replications).

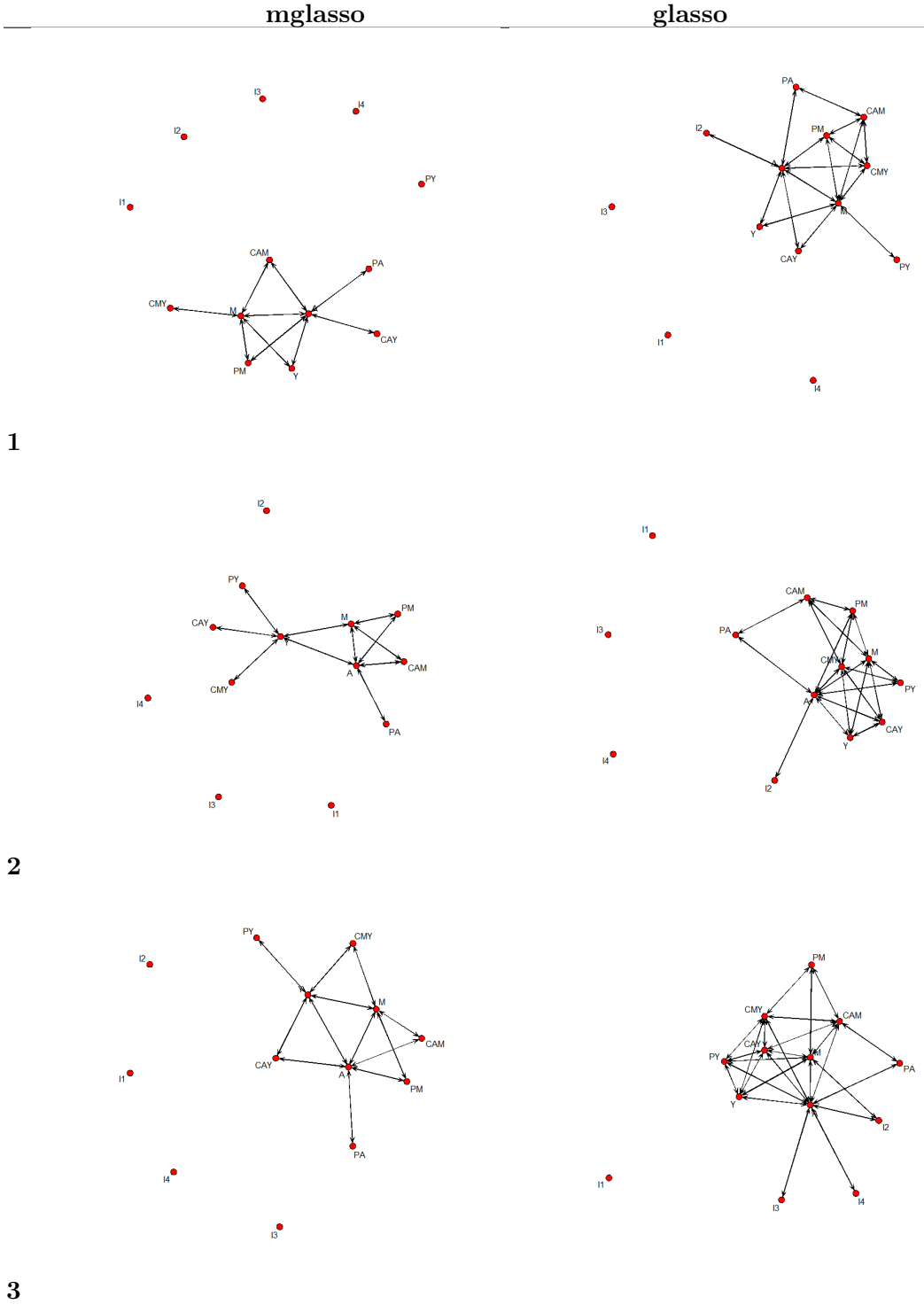
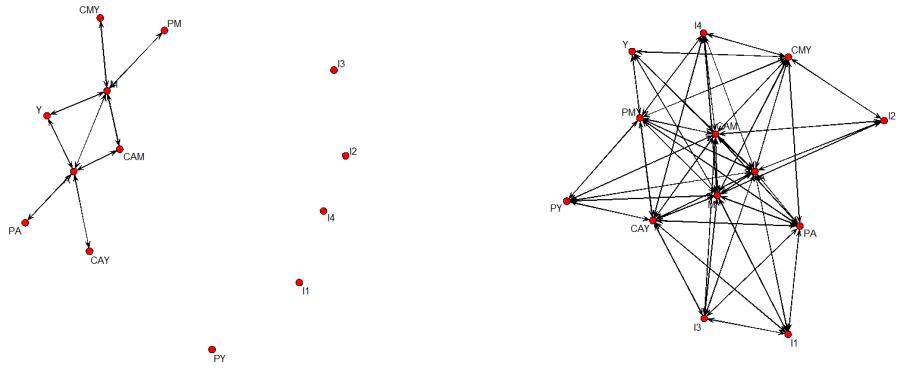
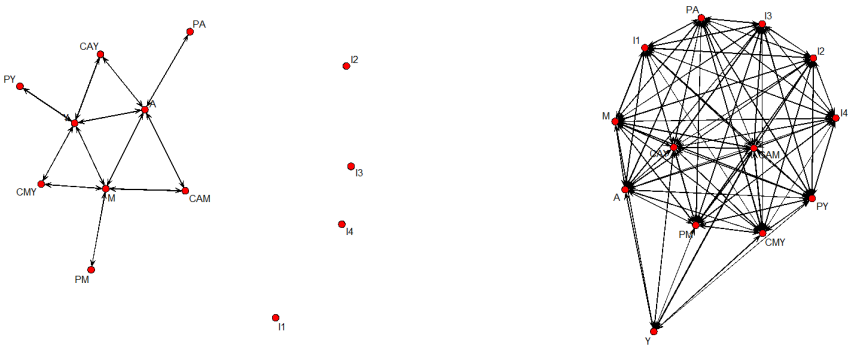


Figure 4.12 Causal mediation network estimation using both mglasso and glasso for binary treatment under Scenarios 1, 2 and 3 (by row) in the low-dimensional setting with $p = 10$ (Case 2) and $n = 500$ (based on 1000 replications).

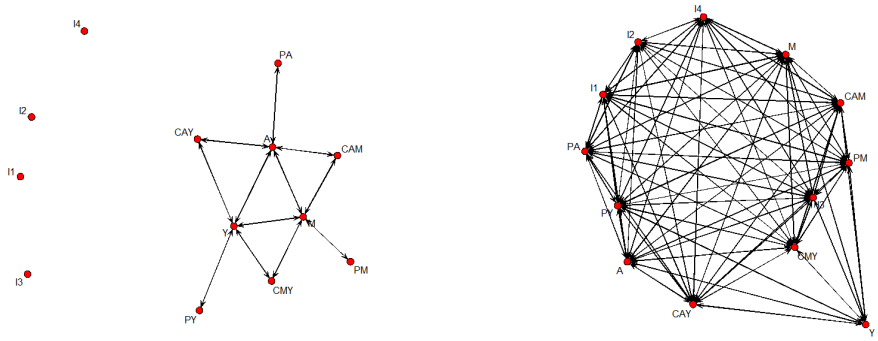
mglasso **glasso**



1



2



3

CONCLUSION

Dans cette thèse de doctorat, nous nous sommes intéressés aux algorithmes d'identification et de sélection des confondants basés sur les données observationnelles et trois différents projets ont été abordés. Nous concluons par une revue des travaux et contributions présentés dans les chapitres précédents. Nous discutons également des extensions possibles de nos travaux. Dans cette thèse nous avons abordé le problème de sélection de confondants en supposant que la vraie structure des données observées n'est que partiellement connue. Tous nos travaux sont basés sur des algorithmes d'apprentissage statistique « data-driven » et ne nécessitent qu'une connaissance partielle a priori de ladite structure de données. Les résultats et contributions présentés dans cette thèse sont très utiles en pratique, car la vraie nature des variables est souvent inconnue. En tout, trois articles scientifiques ont été écrits dans le cadre de cette thèse.

Le premier article « Generalized outcome adaptive lasso (GOAL) » a été publié dans le journal *Biometrics* sous la forme d'une note « réaction du lecteur » de l'article par Shortreed et Ertefaie (2017). Ce travail a été motivé par la contribution importante de Shortreed et Ertefaie (2017) pour la sélection des confondants à travers l'approche « outcome-adaptive lasso (OAL) ». OAL est un algorithme intelligent basé sur le lasso adaptatif qui permet de cibler les variables appropriées à inclure dans le modèle de score de propension et exclure celles qui sont nuisibles. Alors que OAL a montré une bonne performance, en petite dimension et avec des prédicteurs indépendants, sa performance en grande dimension et pour les prédicteurs corrélés est discutable. Pour remédier à ces limites de OAL, nous avons proposé GOAL pour généraliser et améliorer l'approche OAL. GOAL combine le lasso adaptatif et l'elastic net pour résoudre simultanément les problèmes de sparsité et de colinéarité. Les résultats de simulation ont montré que GOAL améliore OAL sur des données de grande dimension ou corrélées.

Il est à noter que nous avons développé notre approche GOAL avec un nombre modeste et important de covariables (p) (lorsque $p < n$). Une extension potentielle de ce travail serait de généraliser GOAL pour le cas $p \geq n$ appelé ultra grande dimension. En effet, en raison des propriétés de l'elastic net adaptatif, GOAL semble bien équipé pour s'attaquer à ce cas. Cependant, GOAL (ainsi que OAL) est basé sur les poids obtenus de la méthode des moindres carrés ordinaires $\hat{w}$ qui nécessitent un modèle de rang complet ($n > p$) pour leur estimation. L'extension de GOAL à l'ultra grande dimension nécessitera donc quelques modifications dans la définition des poids adaptatifs. Ceci sera étudié dans une future étude.

Le deuxième article « Outcome-adaptive regularization for causal mediation analysis (MOAR) » est en voie d'être soumis sous la forme d'un article original dans le journal *Statistics in Medicine*. Ce travail est motivé par les récentes approches prometteuses de sélection de confondants, OAL (Shortreed et Ertefaie, 2017) et GOAL (Baldé *et al.*, 2022), pour l'estimation de l'effet total du traitement sur la réponse. L'analyse de la médiation causale cherche à décomposer l'effet total du traitement sur la réponse d'intérêt en effet direct et indirect. Aujourd'hui, il y a un intérêt croissant des chercheurs à développer des méthodes pour estimer l'effet direct et indirect d'un traitement à travers un médiateur spécifique. Il est alors particulièrement important de développer des méthodes de sélection des confondants dans un contexte de médiation. C'est pourquoi nous avons développé l'algorithme MOAR pour sélectionner les bonnes variables dans un contexte d'analyse de la médiation. Les résultats de simulations ont montré que OAL et GOAL peuvent exclure des covariables importantes comme les confondants médiateur-réponse et inclure des variables nuisibles comme les prédicteurs purs du médiateur, dans l'estimation de l'effet direct. En conséquence, OAL et GOAL présentent un biais pour l'effet direct, alors que notre approche proposée MOAR améliore OAL et GOAL lors de l'estimation de l'effet direct. Globalement, MOAR fonctionne pour tous les effets (effet total, direct et indirect) comme l'estimateur qui inclut uniquement les variables souhaitables pour une estimation efficace et sans biais de ceux-ci.

Nous avons développé MOAR en considérant un nombre relativement petit de covariables

non corrélées. De par sa définition, MOAR pourrait être utilisé dans un contexte de grande dimension ou avec des covariables corrélées. On pourrait vérifier la performance de MOAR pour les problèmes de grande dimension et ultra grande dimension dans une étude future.

Le troisième projet s'intitule « Graphical Lasso networks in the context of causal mediation analysis (mglasso) ». Dans cet article, nous avons développé une nouvelle approche basée sur le type de lasso graphique (glasso) de Friedman, Hastie et Tibshirani (2008) dans le contexte de l'analyse de médiation causale. Le glasso est un algorithme très pratique et largement utilisé pour estimer l'inverse de la matrice de variance-covariance à partir de données normales multivariées observées. Malgré la bonne performance de glasso pour les graphiques non dirigés, sa capacité d'identifier la vraie structure des données de médiation causale est discutable. C'est dans ce contexte que nous avons proposé mglasso pour améliorer glasso pour les données de médiation causale. Les résultats de simulation montrent que mglasso est capable d'identifier la vraie nature des données dans un contexte de médiation et améliore ainsi mglasso.

Une interface graphique simple est en voie de production. Elle pourrait être très utile en pratique et servir d'outil de visualisation des relations directes et indirectes probables entre variables.

RÉFÉRENCES

Acharya, A., Blackwell, M. et Sen, M. (2016). Explaining causal findings without bias: Detecting and assessing direct effects. *American Political Science Review*, *110*(3), 512–529.

Algamal, Z. Y. et Lee, M. H. (2015a). Applying penalized binary logistic regression with correlation based elastic net for variables selection. *Journal of Modern Applied Statistical Methods*, *14*(1), 168–179.

Algamal, Z. Y. et Lee, M. H. (2015b). Regularized logistic regression with adjusted adaptive elastic net for gene selection in high dimensional cancer classification. *Computers in Biology and Medicine*, *67*, 136–145.

Arria, A. M., Caldeira, K., Bugbee, B., Vincent, K. et O'Grady, K. (2015). The academic consequences of marijuana use during college. *Psychology of addictive behaviors : journal of the Society of Psychologists in Addictive Behaviors*, *29*(3), 564–575.

Baldé, I. et Lefebvre, G. (2022+). Outcome-adaptive regularization for causal mediation analysis: Variable selection for direct and indirect effects. *Cet article est sur le point d'être soumis.*

Baldé, I., Yang, Y. A. et Lefebvre, G. (2022). Reader reaction to "Outcome-adaptive lasso: Variable selection for causal inference" by Shortreed and Ertefaie (2017). *Biometrics*, 1–7. https://onlinelibrary.wiley.com/doi/pdf/10.1111/biom.13683.

Banerjee, O., Ghaoui, L.-E. et d'Aspremont, A. (2008). Model selection through sparse maximum likelihood estimation. *Journal of Machine Learning Research*, 9: 485–516.

Baron, R. M. et Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182.

Brookhart, M. A., Schneeweiss, S., Rothman, K., Glynn, R. J., Avorn, J. et Stürmer, T. (2006). Variable selection for propensity score models. *American Journal of Epidemiology*, *163*(12), 1149–1156.

Dahl, J., Vandenberghe, L. et Roychowdhury, V. (2008). Covariance selection for nonchordal graphs via chordal embedding. *Optimization Methods and Software*,

23(4), 501–520.

De Luna, X., Waernbaum, I. et Richardson, T. (2011). Covariate selection for the nonparametric estimation of an average treatment effect. *Biometrika*, *98*(4), 861–875.

Diop, A., Lefebvre, G., Duchaine, C. S., Laurin, D. et Talbot, D. (2021). The impact of adjusting for pure predictors of exposure, mediator, and outcome on the variance of natural direct and indirect effect estimators. *Statistics in Medicine*, *40*(10), 2339–2354.

Efron, B., Hastie, T., Johnstone, I. et Tibshirani, R. (2004). Least angle regression. *Annals of Statistics*, *32*(2), 407–499.

Fan, J. et Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, *96*(456), 1348–1360.

Fan, J. et Li, R. (2006). Statistical challenges with high dimensionality: Feature selection in knowledge discovery. *Proceedings of the International Congress of Mathematicians, European Mathematical Society*, *3*, 595–622.

Fan, J., Liu, H., Ning, Y. et Zou, H. (2017). High dimensional semiparametric latent graphical model for mixed data. *Journal of the Royal Statistical Society : Series B*, *79*(2), 405–421.

Fan, J. et Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society : Series B*, *70*(5), 849–911.

Fan, J. et Peng, H. (2004). Nonconcave penalized likelihood with a diverging number of parameters. *Annals of Statistics*, *32*(3), 928–961.

Farbmacher, H., Huber, M., Lafférs, L., Langen, H. et Spindler, M. (2021). Causal mediation analysis with double machine learning. *https://arxiv.org/abs/2002.12710*.

Friedman, J., Hastie, T. et Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, *9*(3), 432–441.

Friedman, J., Hastie, T. et Tibshirani, R. (2011). glasso: Graphical lasso-estimation of gaussian graphical models. R package version 1.7. *http://CRAN.R-project.org/package=glasso*.

Galloway, M. (2018). Cvglasso: Cross validation package for the popular glasso package. *https://cran.r-project.org/web/packages/CVglasso/index.html*.

Ghosh, S. (2007). Adaptive elastic net: A doubly regularized method for variable selection to achieve oracle properties. *Tech. Rep. pr07-01, available at http://www.math.iupui.edu/research/preprints.php, IUPUI*.

Ghosh, S. (2011). On the grouped selection and model complexity of the adaptive

elastic net. *Statistics and Computing*, *21*, 451–462.

Hernán, M. A. et Robins, J. M. (2020). Causal inference: What if. *Boca Raton: Chapman Hall/CRC*.

Hirano, K. et Imbens, G. W. (2001). Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization. *Health Services and Outcomes Research Methodology*, *2*, 259–278.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*(396), 945–960.

Imai, K., Keele, L. et Tingley, D. (2010a). A general approach to causal mediation analysis. *Psychological Methods*, *15*(4), 309–334.

Imai, K., Keele, L., Tingley, D. et Yamamoto, T. (2010b). Causal mediation analysis using R. in H. D. Vinod (Ed.), advances in social science research using R. *Psychological Methods*, *196*, 129–154.

Imai, K., Keele, L. et Yamamoto, T. (2010c). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science*, *25*(1), 51–71.

Islam, M. S. et Noor-E-Alam, M. (2021). Feature selection for causal inference from high dimensional observational data with outcome adaptive elastic net. *https://arxiv.org/abs/2111.13800*.

Jones, J., Ertefaie, A. et Strawderman, R. L. (2021). Causal mediation analysis: Selection with asymptotically valid inference. *https://arxiv.org/pdf/2110.06127.pdf*.

Kalisch, M., Fellinghauer, B., Grill, E. et al. (2010). Understanding human functioning using graphical models. *BMC Medical Research Methodology*, *10*(14), 1471–2288.

Koch, B., Vock, D. M. et Wolfson, J. (2018). Covariate selection with group lasso and doubly robust estimation of causal effects. *Biometrics*, *74*(1), 8–17.

Lange, T., Vansteelandt, S. et Bekaert, M. (2012). A simple unified approach for estimating natural direct and indirect effects. *American Journal of Epidemiology*, *176*(3), 190–195.

Lange, T., W, H. K., Sørensen, R. et Galatius, S. (2017). Applied mediation analyses: a review and tutorial. *Epidemiol Health. doi: 10.4178/epih.e2017035. PMID: 29121709; PMCID: PMC5723912*, *39*.

Lederer, J. (2022). Graphical models. In: Fundamentals of high-dimensional statistics. *Springer Texts in Statistics. Springer, Cham*.

Li, J., Dong, X., Li, X. et Li, W. (2010). Oracle properties of the adaptive elastic net. *IEEE International Conference on Intelligent Computing and Intelligent Systems*, *3*, 538–542.

Lindmark, A. (2021). Sensitivity analysis for unobserved confounding in causal mediation analysis allowing for effect modification, censoring and truncation. *Statistical Methods and Applications. https://doi.org/10.1007/s10260-021-00611-4*.

Lunceford, J. K. et Davidian, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Statistics in Medicine*, *23*(19), 2937–2960.

Meier, M. H., Hill, M. L., Small, P. J. et Luthar, S. S. (2015). Associations of adolescent cannabis use with academic performance and mental health: A longitudinal study of upper middle class youth. *Drug and alcohol dependence*, *156*, 207–2012.

Meinshausen, N. et Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *Annals of Statistics*, *34*(33), 1436–1462.

Meinshausen, N. et Peter Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society : Series B*, *72*(4), 417–473.

Neyman, J. (1923). On the application of probability theory to agricultural experiments. essay on principles, section 9 (traduit et édité par D. M., Dabrowska et T. P., Speed à partir du texte polonais paru en 1923). *Statistics in Medicine*, *5*(4), 465–472.

Nguyen, T. Q., Schmid, I. et Stuart, E. A. (2021). Clarifying causal mediation analysis for the applied researcher: Defining effects based on what we want to learn. *Psychological Methods*, *26*(2), 255–271.

NIH (2020). Marijuana use at historic high among college-aged adults in 2020. *National Institutes of Health (NIH). https://www.nih.gov/news-events/news-releases/marijuana-use-historic-high-among-college-aged-adults-2020. Wednesday, September 8, 2021.*

Patrick, A. R., Schneeweiss, S., Brookhart, M. A., Glynn, R. J., Rothman, K. J., Avorn, J. et Stürmer, T. (2011). The implications of propensity score variable selection strategies in pharmacoepidemiology: an empirical illustration. *Pharmacoepidemiology and Drug Safety*, *20*(6), 551–559.

Pearl, J. (2009). Causality : Models, reasoning, and inference. 2nd edition. *New York : Cambridge University Press*.

Robins, J. et Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Journal of Personality and Social Psychology*, *3*(2), 143–155.

Robins, J. M. (1997). Causal inference from complex longitudinal data. in: Berkane m. (eds) latent variable modeling and applications to causality. *Lecture Notes in Statistics, vol 120. Springer, New York, NY.*

https://doi.org/10.1007/978-1-4612-1842-5_4.

Rosenbaum, P. R. et Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688–701.

Rubin, D. B. (1997). Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine*, *127*(8 Part 2), 757–763.

Ryan, O., Bringmann, L. F. et Schuurman, N. K. (2022). The challenge of generating causal hypotheses using network models. *Structural Equation Modeling: A Multidisciplinary Journal, DOI: 10.1080/10705511.2022.2056039*, *0*(0), 1–8.

Samoilenko, M. et Lefebvre, G. (2021). Parametric regression-based causal mediation analysis of binary outcomes and binary mediators: moving beyond the rareness or commonness of the outcome. *American Journal of Epidemiology*, *190*(9), 1846–1858.

Shortreed, S. et Ertefaie, A. (2017). Outcome-adaptive lasso: Variable selection for causal inference. *Biometrics*, *73*(4), 1111–1122.

Spirtes, P., Glymour, C. et Scheines, R. (2000). Causation, prediction and search. *2nd ed. Cambridge, MA: MIT Press.*

Steen, J., Loeys, T., Moerkerke, B. et Vansteelandt, S. (2017). medflex : An R package for flexible mediation analysis using natural effect models. *Journal of Statistical Software*, *76*(11), 1–6.

Strobl, R., Grill, E. et Mansmann, U. (2012). Graphical modeling of binary data using the lasso: a simulation study. *BMC Medical Research Methodology*, *12*(16).

Textor, J., Zander, B. V. D., Gilthorpe, M. K., Liskiewicz, M. et Ellison, G. T. H. (2016). Robust causal inference using directed acyclic graphs: the R package 'dagitty'. *International Journal of Epidemiology*, *45*(6), 1887–1894.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*, *58*(1), 267–288.

Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society Series B*, *73*(3), 273–282.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L. et Imai, K. (2014). Mediation : R package for causal mediation analysis. *Journal of Statistical Software*, *59*(5), 1–38.

Uhler, C. (2017). Gaussian graphical models: An algebraic and geometric perspective. *https://arxiv.org/abs/1707.04345.*

Ulbricht, J. (2010). Variable selection in generalized linear models. *Ph.D. Thesis.*

Ludwig Maximilians University Munich.

Valeri, L. et VanderWeele, T. J. (2013). Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. *Psychological Methods*, *18*(2), 137–150.

VanderWeele, T. J. (2015). Explanation in causal inference: Methods for mediation and interaction. *Oxford University Press.*

VanderWeele, T. J. (2018). Short course on causal mediation analysis part I. *Tyler VanderWeele's Faculty Website*, https://www.youtube.com/watch?v=EI5y6pV87–Q.

VanderWeele, T. J. et Vansteelandt, S. (2009). Conceptual issues concerning mediation, interventions and composition. *Statistics and Its Interface*, *2*, 457–468.

VanderWeele, T. J. et Vansteelandt, S. (2010). Odds ratios for mediation analysis for a dichotomous outcome. *American Journal of Epidemiology*, *172*(12), 1339–1348.

Vansteelandt, S. (2009). Estimating direct effects in cohort and case-control studies. *Epidemiology*, *20*(6), 851–860.

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, *20*(7), 557–585.

Ye, Z., Zhu, Y. et Coffman, D. L. (2021). Variable selection for causal mediation analysis using lasso-based methods. *Statistical Methods in Medical Research*, *30*(6), 1413–1427.

Yuan, M. et Lin, Y. (2008). Model selection and estimation in the Gaussian graphical model. *Biometrika*, *94*(1), 19–35.

Zhu, Y. (2021). A brief introduction to causal inference. *First CANSSI-NISS Health Data Science Workshop*, https://www.niss.org/first–canssi–niss–health–data–science–workshop–full–program.

Zierer, J., Pallister, T., Tsai, P. C. et al. (2016). Exploring the molecular basis of age-related disease comorbidities using a multi-omics graphical model. *Scientific Reports*, *6*(37646), https://doi.org/10.1038/srep37646.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, *101*(476), 1418–1429.

Zou, H. et Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B*, *67*(2), 301–320.

Zou, H. et Zhang, H. H. (2009). On the adaptive elastic-net with a diverging number of parameters. *Annals of Statistics*, *37*(4), 1733–1751.