A Graphical Approach to Identifying Structural Bias Using Directed Acyclic Graphs: Its Application to Two-Wave Nonequivalent Control Group Designs

Jaehyun Shin

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In

Educational Research and Evaluation

Yasuo Miyazaki, Chair

David Kniola

Brett D. Jones

Kazuki Hori

February 16, 2026

Blacksburg, Virginia

Keywords: ANCOVA, change scores, bias, causal graphs, Frisch–Waugh–Lovell theorem

A Graphical Approach to Identifying Structural Bias Using Directed Acyclic Graphs: Its Application to Two-Wave Nonequivalent Control Group Designs

Jaehyun Shin

ABSTRACT

It is well known that the analysis of covariance (ANCOVA) and the change-score analysis (CSA) can produce quite different treatment-effect estimates when applied to data from two-wave nonequivalent control-group designs, a phenomenon known as the Lord's paradox. Pearl's (2009) structural causal model (SCM) provides a useful and intuitively appealing tool to address the Lord's paradox. Using the SCM, Kim and Steiner (2021) combined the backdoor criterion with the path-tracing rules and showed that it identified the exact bias for the CSA. Though they implied that this graphical causal model approach could be applied to the ANCOVA case in a similar way, they did not explicitly show the details. Therefore, in the present study, to examine their implication, I applied the graphical approach to the ANCOVA and compared the results with the bias derived by the population ordinary least squares (OLS) method (Lüdtke and Robitzsch, 2025). The comparison exhibited a discrepancy, though the core part of the bias obtained by the graphical approach was correct. Specifically, the discrepancy occurred in the terms that were proportional to the core part of the bias implied by each backdoor path. This means that, though the detection of the sources of bias and the identification of the conditions to eliminate the bias could be completed by the graphical approach, the exact quantification of the bias was not possible. To resolve this shortcoming, I applied the so-called regression anatomy formula, also known as the Frisch–Waugh–Lovell (FWL) theorem in econometrics, and found that the proportional term could

be expressed as the residualization-induced scaling factor. I then extended this graphical approach to different data-generating scenarios within two-wave nonequivalent control-group designs and confirmed that it worked well in all cases. The residualization procedure makes a graphical approach self-contained to identify the exact structural bias.

A Graphical Approach to Identifying Structural Bias Using Directed Acyclic Graphs: Its Application to Two-Wave Nonequivalent Control Group Designs

Jaehyun Shin

GENERAL AUDIENCE ABSTRACT

In studies that compare two groups before and after a program, researchers want to know whether the program truly caused any improvement. This question is especially difficult when participants are not assigned to groups at random, because the groups may already differ in ways that affect the outcome. In these studies, two common statistical methods can give different answers even when they are applied to the same data: one method adjusts the later score for the earlier score (ANCOVA), and the other looks directly at how much scores change over time (change-score analysis). This disagreement is sometimes called Lord's paradox. This study examined why that disagreement happens. I used simple causal diagrams to trace how preexisting differences between groups can create systematic error, or bias, in the estimated effect of a program. The diagrams showed where the bias came from, but they did not fully show how large that bias would be. I then used a result from the Frisch–Waugh–Lovell (FWL) theorem to identify the missing piece: a factor that determines how much each source of bias affects the final estimate. Adding that factor completed the diagram-based explanation and made it possible to recover the exact bias in ANCOVA. I also tested the results across several realistic scenarios and found that they matched the bias obtained from ordinary least squares (OLS) method in every case. Overall, this study showed more clearly when ANCOVA and change-score analysis can be trusted and when they can mislead researchers in before-and-after studies without random assignment.

ACKNOWLEDGEMENTS

I am sincerely grateful to my advisor, Dr. Yasuo Miyazaki, for his guidance throughout my doctoral journey. He taught me not only how to conduct research, but also how to approach it with discipline, integrity, and independence. I am grateful to Dr. David Kniola for his encouragement, kindness, and the opportunities that sustained me through this process. I also thank Dr. Brett D. Jones and Dr. Kazuki Hori for their thoughtful advice in bringing this dissertation to completion.

I owe much to my parents, Dongwoo Shin and Yeonsun Lim, and my parents-in-law, Youngil Ko and Jeongha Kim. Their love, faith in me, unwavering support, and prayers have carried me to this point. They have always been role models for Hayoung and me. In moments of exhaustion, I drew strength from their example and the quiet steadiness of their care. The respect I have for them, and my gratitude for all they have given me, go beyond what words can convey.

Above all, I owe more than words can express to my beloved wife, Hayoung Ko. Marrying you has been the greatest blessing of my life. You cared for me through every season of hardship and rejoiced in each moment of joy more than I did. When I doubted myself, you believed in me more firmly than I believed in myself, and that gave me the courage to keep going. You celebrated every accomplishment of mine and always made me feel appreciated, even for the ordinary things I did as your husband. Because of your love and belief, I was able to complete this degree with hope, and with you beside me, I feel ready to face whatever lies ahead.

Finally, to our beloved son, Ian Shin, your arrival has filled our lives with a joy we had never known before. From the moment we knew you were coming, your mother and I began to know a happiness deeper than we could ever have imagined, and after you were born, that joy only deepened. Your mother and I are deeply grateful for the gift of your life, and we pray that you will always grow in health, happiness, and love.

Table of Contents

1. Introduction

Obtaining an unbiased causal effect of interventions or policies on outcomes of interest in the presence of unobserved confounders remains a central goal in evaluation research across the social sciences. Whether the focus is on student achievement in education, therapeutic outcomes in psychology, or organizational performance in management, the fundamental question revolves around understanding whether a given treatment truly causes an observed improvement (or decline) in the targeted outcome. When randomization is not feasible, researchers frequently employ the two-wave (i.e., pretest and posttest) nonequivalent control group design, in which researchers administer an intervention to one group while a non-randomly assigned comparison group does not receive the intervention (Cook & Campbell, 1979, p. 103; Shadish et al., 2002, p. 136; Whitley & Kite, 2013, p. 290). When randomization is not feasible, one of the most frequently employed quasi-experimental methods is the two-wave (pretest and posttest) nonequivalent control group design, in which a treatment or intervention is administered to one group while a non-randomly assigned comparison group does not receive the treatment or the intervention (Cook & Campbell, 1979, p. 103; Shadish et al., 2002, p. 136; Whitley & Kite, 2013, p. 290).

The classical criteria proposed by the nineteenth-century philosopher John Stuart Mill (1843/2023) still helps researchers to find the fundamental logic of causal relationships in two-wave nonequivalent control group design. This logic requires temporal precedence (i.e., the treatment comes before the posttest), statistical association (i.e., the treatment and the outcomes covariate), and the elimination of alternative explanations (i.e., no confounder accounts for the observed relationship) (Cook & Campbell, 1979, p. 18; Kenny, 1979, p. 3; Bollen, 1989, pp. 40-

67[1]; Shadish et al., 2002, p. 6.). The two-wave nonequivalent control group design enables researchers to examine the changes in outcome variables over time and compare the treatment effect between two nonequivalent groups. Thus, this research design, a common approach in quasi-experimental research, can effectively meet the first two of the principles for establishing causality but often struggles with the third due to the inherent nonequivalence of the groups involved. Therefore, the major challenge for the two-wave nonequivalent control group design lies in meeting the third criterion – eliminating alternative explanations. Due to the nonequivalent nature of the groups, there can be pre-existing differences between the treatment and control groups that might account for the observed effects rather than the treatment itself. These pre-existing differences make it difficult to confidently assert that the treatment is the sole cause of any observed changes. Without random assignment, it is challenging to ensure that the two groups are comparable in all respects except for the treatment, leaving room for alternative explanations for any observed differences in outcomes.

Two principal statistical methods, Analysis of Covariance (ANCOVA) and change score analysis, are often used to address baseline differences in two-wave nonequivalent control group designs (Allison, 1990; Maxwell et al., 2018, pp. 498–502). Although both methods incorporate some form of adjustment for the pretest measurement, they can, under certain circumstances, yield contradictory conclusions about whether the treatment has an effect. This quandary was famously highlighted by Lord's paradox (Lord, 1967, 1969, 1975), in which the same dataset was analyzed by two different methods that led to opposing interpretations about the treatment's impact. At the core of Lord's paradox lies the fact that regression-based analytic methods, such as ANCOVA or

[1] Although Bollen (1989) did not mention John Stuart Mill in chapter three of his book, the three components for the definition of cause: the direction of influence, association, and isolation are like Mill's classical criteria for the logic of causal relationships.

change score analysis, do not by themselves clarify the underlying structure of causal relationships among variables. Instead, they operate at the level of associations, attempting to statistically "control" or "adjust" for the influence of one variable on another without making explicit what noncausal paths might exist or whether a variable is a confounder, mediator, or collider.

Recognizing the limitations of purely regression-based methods, the graphical causal model (Pearl, 2009; Pearl, 2016) is particularly noteworthy for its ability to visually represent hypothesized causal structures among variables (i.e., data-generating model) using directed acyclic graphs (DAGs). By drawing arrows to denote potential causal connections, researchers can distinguish between direct causal paths, indirect causal paths, and noncausal paths. This visual approach helps clarify whether a variable is an antecedent that confounds the relationship between treatment and outcome or whether it plays a mediating or moderating role. The backdoor criterion, a key principle in the graphical causal model, dictates that adjusting for a set of variables that blocks all backdoor paths from treatment to outcome is sufficient to obtain the unbiased causal effect. In other words, once all noncausal routes carrying spurious associations are closed, the path that remains captures the true treatment effect.

In this dissertation, I define the term treatment effect in the context of graphical causal models as the difference in an outcome variable under two hypothetical conditions: one in which the unit is conceptually assigned to receive the treatment and one in which it is not, with all other factors held constant. This aligns with the *ceteris paribus* condition (Wooldridge, 2020, p. 67). Within the DAGs framework, any non-causal paths are regarded as "blocked," whether this occurs through the structure of the DAG itself (e.g., a collider that is not conditioned on) or through statistical conditioning on relevant variables. Under these conditions, the difference in the outcome can be attributed solely to the treatment, rather than to other factors that might otherwise influence

the relationship.[2] Throughout this dissertation, I therefore define the treatment effect as the direct effect of treatment on outcome. Any influence of treatment on outcome that does not operate through this direct path, even when it follows a genuinely causal path (for example through a mediator), is treated as structural bias rather than as part of the treatment effect.

When applied to a two-wave nonequivalent control group design, the graphical approach, which combines the path-tracing rules with the backdoor criterion (Kim & Steiner, 2021; Pearl, 2013), can be particularly valuable because it visually clarifies how baseline differences, temporal ordering, and confounding variables influence the treatment effect. By mapping out the relationships among the observed variables like the pretest, the posttest, and the treatment assignment, as well as other unobserved variables, researchers can see which assumptions are needed to separate the causal effect of treatment from other influences. For instance, baseline group differences may be correlated with both treatment assignment and posttest scores, creating a backdoor path that biases treatment effect. Visualizing such paths is often more accessible than relying on algebraic population Ordinary Least Squares (OLS), which requires parsing covariance and variance terms in regression equations. Although the OLS is precise in identifying[3] structural bias, it can be less intuitive and does not make it explicit that unobserved variables may drive bias. In ANCOVA as a multiple-regression model, applying the Frisch–Waugh–Lovell theorem (i.e., regression anatomy) decomposition provides a natural algebraic complement to the graphical approach because it expresses the treatment coefficient, after partialling out the pretest, as a simple

[2] Here, I used the causal effect and treatment effect interchangeably since the context that we consider is the causal effect that includes the treatment conditions created by researchers. In this sense, the term causal effect is used in a narrow sense to indicate that the treatment manipulated by the researcher existed. In this dissertation, I will keep using the terms this way (i.e., treatment effect and the causal effect interchangeably).

[3] In this dissertation, I use the verb "identify" in a broad, nontechnical sense when discussing structural bias, meaning to derive or make explicit the population bias implied by an assumed data-generating model. When discussing the treatment-effect parameter, I use "identification" in the formal statistical sense, meaning that the parameter is uniquely determined by the observed data distribution under the assumed model.

residual-covariance ratio; this makes the role of adjustment transparent and enables a direct term-by-term comparison with the structural bias identified by the graphical approach

The existing literature, however, has not yet shown in a single framework how these two approaches relate in two-wave nonequivalent control group designs. Kim and Steiner (2021) used DAGs to leverage the advantages noted above, but their analysis centered on the change scores and did not extend the graphical approach to ANCOVA. While Lüdtke and Robitzsch (2025) compared the structural bias between ANCOVA and change score analysis, they did not employ DAG tools but relied on the OLS derivations only. Thus, this dissertation provides a side-by-side comparison of the structural bias identified by the graphical approach and that identified by the OLS, and it states the corresponding unbiasedness conditions across a set of scenarios distinguished by the number of additional unobserved confounders (beyond the pretest) in two-wave nonequivalent control group designs. Within each scenario, the subsequent cases (i.e., data-generating model) are organized by a pretest-as-mixture premise, implemented by varying the assumed pretest-to-posttest relation. Under this premise, a central question is whether, and when, a pretest has a causal effect on a posttest. Prior work often includes a direct pretest-to-posttest arrow because the pretest predicts the posttest, but prediction alone is not a causal rationale. Allison (1990) offered a helpful stock versus flow distinction, yet real test scores rarely represent a pure stock or a pure flow. I therefore adopt a pretest-as-mixture perspective: pretests blend stock-like components that persist (for example, durable knowledge and practiced test-taking proficiency) with flow-like components that are newly produced and can be reinforced (for example, motivational factors after knowing the pretest score). This perspective explains when a direct pretest-to-posttest effect should remain even with an ability confounder present and when theory

instead posits a mediating path that captures a between-wave change process such as motivational factors (for example, self-efficacy).

To this end, this dissertation makes two contributions. First, it compares structural bias and sufficient unbiasedness conditions for ANCOVA and change-score analysis across two-wave scenarios using graphical approach and population-level OLS. Second, it completes graphical approach for the structural bias of the population-level ANCOVA estimator by making residualization explicit through the regression anatomy formula. Taken together, these contributions show that the path-tracing ruels combined with the backdoor criterion provides an intuitive and accessible approach to identifying structural bias in two-wave designs, particularly relative to population-level OLS, which relies on a more algebraic characterization. By linking the graphical and algebraic results, this dissertation clarifies the causal assumptions under which ANCOVA and change-score analysis recover the population treatment effect in two-wave nonequivalent control-group designs and offers applied researchers a transparent way to reason about bias without relying exclusively on algebraic derivations.

2. Literature Review

This literature review builds the conceptual foundation for the dissertation's side-by-side bias analysis of ANCOVA and change score methods in two-wave nonequivalent control group designs. I begin by revisiting Lord's paradox and the traditional regression-based approaches used to address baseline nonequivalence, with focused reviews of ANCOVA and change score analysis and what the existing literature concludes about their performance and limitations. I then introduce graphical causal models emphasizing how DAGs can be used to diagnose structural sources of bias that the analytic method alone does not reveal. Finally, I motivate the dissertation's scenario and case structure through the "pretest as a mixture of stocks and flows" perspective, which provides a causal rationale for when a pretest–posttest relationship is represented as absent, direct, or mediated in the data-generating model.

2.1. Traditional Methods to Lord's Paradox

Four specific examples were employed to introduce the Lord's paradox across Psychological Bulletins (Lord, 1967, 1969) and a book chapter in the Encyclopedia of Educational Evaluation (Lord, 1975, p. 235). Although each example efficiently illustrated how the two statistical analytic methods for the same data produce different results, this has been another reason why this paradox is considered more complex since researchers have focused on different examples in discussions surrounding Lord's paradox so far. Among the different specific examples of Lord's paradox, the Head Start program example is selected; to focus on in this dissertation and see why it conforms to the two-wave nonequivalent control group design. Then two analytic methods: Analysis of Covariance (ANCOVA) and the change score analysis are reviewed, which lie at the heart of the paradox as traditional linear models, before applying the causal inference

framework. When examining the same data, ANCOVA estimates the treatment effect on the posttest outcome while adjusting for the pretest score as a covariate, whereas change score analysis estimates the treatment effect on the pre–post difference (i.e., the change score) as the outcome, and the two methods can sometimes yield differing conclusions (Wright, 2006). The ANCOVA and change score analysis methods have utilized a linear regression model to comprehend the Lord's paradox in this dissertation.

2.1.1. Two-wave Nonequivalent Control Group Design in Lord's Paradox

Maxwell et al. (2018, p. 479) articulated the rationale of the quasi-experimental designs: "Many events of profound consequence for human behavior (e.g., trauma or disease, changes in public policy, or social innovations) cannot feasibly or ethically be manipulated in a randomized experiment. In such cases, the only feasible way of quantitatively estimating the effect of the natural event or treatment experienced by a group is often by comparison to another, non-randomized group that did not experience the event or treatment." This only feasible way is known as quasi-experimental designs, and the two-wave nonequivalent control group design with pretest and posttest, is perhaps the most frequently used quasi-experimental design where the random assignment is not feasible due to ethical and practical reasons (Cook & Campbell, 1979, p. 103; Shadish et al., 2002, p. 136; West et al., 2000). This design is primarily characterized by its inclusion of two groups, typically a treatment group and a control group, which are not created through random assignment but are observed over two time points - before (pretest) and after (posttest) the implementation of a treatment or intervention. Cook and Campbell (1979) initially presented this design as a practical alternative to randomized controlled trials, particularly in contexts where randomization is either impractical or impossible, as this design satisfies John Stuart Mill's classical criteria for the logic of causal relationships despite the potential violation of

the third criterion: the elimination of alternative explanations (i.e., no confounder accounts for the observed relationship). Shadish et al. (2002) further emphasized the significance of this design in drawing causal inferences in natural settings while also acknowledging the potential threats to internal validity of the treatment effect due to the absence of random assignment.

In the two-wave nonequivalent control group design (e.g., untreated control group design with dependent pretest and posttest), several key variables are diagrammed:

NR X T Y

NR X Y

Both groups undergo a pretest (X) to establish baseline measurements, but only the treatment group receives an intervention (T), followed by a posttest (Y) for both groups. Because the groups are nonequivalent due to the absence of random assignment (NR), systematic bias is presumed to be present. For example, unobserved confounders that make pre-existing differences between the groups such as ability, motivation, or demographics could influence the outcomes observed in the posttest. These pre-existing differences between the groups in the nonequivalence control group design lie in the nature of the Lord's paradox, a phenomenon where differing analytic methods can lead to contrasting conclusions. This paradox highlights the difficulty in ascertaining whether observed changes in the posttest are attributable to the treatment or to pre-existing differences between the groups. The paradox underscores the complexity of interpreting results in non-randomized settings, where the same data can suggest different inferences about the treatment's effectiveness depending on the method of analysis.

The exploration of Lord's paradox through various exemplars has yielded significant insights into the interpretation of group comparisons with the pre-existing differences, as chronicled in Lord's foundational works in Psychological Bulletin (Lord, 1967, 1969) and further

encapsulated in the Encyclopedia of Educational Evaluation (Lord, 1975, p. 235). To select an example of Lord's paradox that best matches the two-wave nonequivalent control group design targeted in this dissertation, it is necessary to determine the variables that can be defined within it accurately. I looked closely at the original example covered in Fredrick Lord's (1967) two-page article to extract the variables that could consist of the two-wave nonequivalent control group design.

> A large university is interested in investigating the effects on the students of the diet provided in the university dining halls and any sex difference in these effects. Various types of data are gathered. In particular, the weight of each student at the time of his arrival in September and his weight the following June are recorded (p.304).

Lord (1967)'s original example, however, fell short of satisfying the criteria for causal analysis as outlined by Rubin's principle of "no causation without manipulation" (Rubin, 1975; Rubin, 2008). Gender, being an unmanipulable characteristic, precluded the use of this example for causal inference as it did not allow for the experimental manipulation of the treatment. In that sense, the other way was to divide the students into two groups based on the two different dining halls, A and B, where the treatment of interest is the effect of diet, not of gender.

Wainer and Brown (2004) illustrate Lord's paradox using a simple dining-room example that frames the dining room assignment as an intervention on students' diet. This paradox arises because two common analytic strategies can lead to different conclusions about the same two-wave data. Suppose a school has two dining rooms, A and B, and students' weights are measured at two time points, in September (pretest) and in the following June (posttest). The goal is to assess whether eating in one dining room versus the other is associated with differential weight gain over the school year.

In this setting, a change score analysis focuses on the average within-student change, comparing $Y - X$ across dining rooms. Under the data configuration discussed by Wainer and Brown (2004), this approach suggests little to no difference in mean weight change between dining rooms. An ANCOVA, in contrast, compares dining rooms on the posttest outcome while statistically holding constant the September baseline weight. Because posttest weight is typically strongly related to baseline weight, this adjustment effectively compares Dining Room A and B at a common starting weight. When the two dining rooms differ in their baseline weight distributions, this "equal baseline" comparison can yield a nonzero difference in the adjusted June weights even when the raw average changes appear similar. Thus, the apparent paradox reflects the fact that change score analysis and ANCOVA target different population comparisons: the former contrasts mean observed change in the two groups as they are, whereas the latter contrasts expected posttest outcomes for students who share the same baseline weight.

The Head Start program, launched in 1965 and still operated by the Office of Head Start within the U.S. Department of Health and Human Services, provided an example that has all the information to describe the two-wave nonequivalent control group design. The Head Start program is a federal initiative in the United States designed to provide early childhood education, health, nutrition, and parent involvement services specifically to low-income children from birth to age five and their families, with the primary aim of preparing these children for school by enhancing their cognitive, social, and emotional development. Within the context of the Head Start program example, the Head Start program could be defined as the "treatment" when estimating the effect of the educational intervention program. That is, the term "treatment effect" refers to the quantifiable impact that the Head Start program had on the outcome of interest to the participating children compared to those who did not participate. If the outcome of interest is academic

achievement, then intuitively, the treatment effect of the Head Start program could be estimated as the average increase or decrease of the academic achievement test scores in the treatment group compared to the control group if certain assumptions are met.

The first major evaluation of the Head Start program was conducted in 1969 (Cicirelli, 1969) then Campbell and Erlebacher (1970) indicated this phenomenon in their research followed by Lord (1975) to present this as an actual example of his paradox.

> This paradox is not just an amusing statistical puzzle. Its purpose is to call attention to a common flaw in certain types of research. For example, consider the problem of evaluating federally funded special educational programs. A group of disadvantaged children are pretested in September, then enrolled in a special program, and finally posttested in June. A control group of children are similarly pretested and posttested but not enrolled in the special program (p.235).

The Head Start program example is characterized by its explicit two-level manipulable treatment (i.e., Head Start group vs. untreated control group) with the outcomes for both groups measured in the pretest and the posttest. Here, the treatment group is represented by the children who participated in the Head Start program, and the control group by those who did not, allowing for a clear delineation in treatment levels. This untreated control group allows for a more explicit evaluation of the treatment effect by comparing outcomes against those who participated in the program. Holland and Rubin (1983) recognize this as the two explicit treatments, where the manipulation of the intervention (i.e., children receiving the Head Start program versus those who do not) is explicit. Furthermore, it is reasonable to assume that the students were assigned to a treated (i.e., Head Start group) or untreated group (i.e., control group) based on the pretest results so that it can be considered a nonequivalent control group design in which random assignment was

not implemented. With its manipulable treatment, two-wave design, and nonequivalent control group where the random assignment was inappropriate, this Head Start program example is the most suitable case study for applying the causal inference frameworks for Lord's paradox. Hong (2015) further leverages this example to apply Rubin's causal model or the potential outcomes model, elucidating Lord's paradox, and the pursuit of an unbiased treatment effect by delineating the necessary assumptions for each analytic method. In this dissertation, I use the term "Lord's paradox" to refer to the statistical phenomenon described in the Head Start program example where the ANCOVA and change score analysis methods yielded the different conclusion with the same data from a two-wave nonequivalent control group design.

2.1.2. Analysis of Covariance (ANCOVA)

Fisher (1970), who developed the ANCOVA technique as well as the ANOVA, originally considered that the ANCOVA will increase the precision (i.e., reduction of within-group error variation associated with the covariates) of the estimate in identifying the treatment effects by adjusting for the influence of concomitant variables (i.e., covariates) in the context of the design of experiments which utilizes a random assignment, and indicated that "it combines the advantages and reconciles the requirements of the two very widely applicable procedures known as regression analysis and analysis of variance" (p. 283). Therefore, ANCOVA was originally developed as an extended version of ANOVA that focuses on manipulable group factors that can be randomly assigned rather than non-manipulable group factors that cannot be randomly assigned to the groups. However, the researchers have paid attention to another advantage of ANCOVA besides the increase of statistical power in experimental design. According to Cohen et al. (2003, p. 350)'s definition, "Analysis of Covariance (ANCOVA) is an analysis strategy typically applied to assess the impact of one (or more) group factors (e.g., treatment groups; gender) while statistically

controlling for other IVs (called the covariates)". Here, we know that Cohen et al. (2003) gave both types of group factors manipulable (e.g., treatment group) in terms of (quasi-) experimental study and non-manipulable (e.g., gender) factors in observational study. In other words, ANCOVA is regarded as a special case of the general linear model that accounts for differences on the covariate between groups by incorporating it as a continuous predictor variable (Maxwell et al., 2018, p. 453). Specifically, Pedhazur (1997, p. 653) clarified the advantage of ANCOVA in that it can make originally nonequivalent groups equivalent on the observed covariates when the random assignment is not feasible in social or educational intervention program. Since this advantage of ANCOVA to equate groups that are essentially nonequivalent, it becomes particularly useful for two-wave design to compare the post-treatment outcomes when preexisting differences (i.e., pretest score differences) exist between treated and untreated groups such as quasi-experimental research like the Head Start program. For example, in the Head Start program (Lord, 1975), the Head Start group had a higher proportion of disadvantaged children and, as a result, had lower pretest average scores than the control group. ANCOVA could adjust for this inequivalence in the pretest scores so that the difference in the posttest is attributed to the treatment effect.

The regression model of the ANCOVA for the two-wave data scenario can be defined as follows:

$$Y = \beta_0 + \gamma_{\text{ANCOVA}}\, T + \beta_1\, X + \varepsilon_1 \qquad (1)$$

where the dichotomous treatment variable T is regressed on the posttest score Y while controlling for the single observed covariate, the pretest score, X. The β_0 is the regression intercept, and the ANCOVA estimator γ_{ANCOVA} and β_1 are the regression coefficients of the regression of Y on T and X respectively. The ANCOVA regression model described in Equation (1), where the treatment

effect with a specific coefficient, could offer a direct measure of its impact on the outcome variable while controlling the baseline (i.e., pretest) scores. By incorporating the pretest score as a covariate, the ANCOVA analytic model statistically adjusts for the baseline differences, ensuring that the observed differences in posttest scores are attributed to the treatment rather than preexisting or baseline differences. A statistically significant ANCOVA estimator implies that there is a group difference in the posttest scores after adjusting for pretest scores. This means that ANCOVA controls initial differences in pretest scores between the groups, as if both groups had started with equivalent pretest scores. The standard assumptions in multiple regression analysis and the unique assumption under ANCOVA should be met. The internal validity of the treatment effect by ANCOVA relies on the unique assumption, the homogeneity of regression slopes, besides the standard assumptions of the error term, ε_1, in the context of the regression analysis (e.g., homoscedasticity, normality, independence of the errors, zero conditional mean of the error assumption,[4] and no measurement error for the measured variable especially for the independent variables). The homogeneity of regression slopes necessitates that the effect of the covariate on the dependent variable remains consistent across different groups which means that there is no interaction between the treatment and the covariate.

Note that these assumptions of the traditional ANCOVA as a statistical analytic model are primarily for establishing unbiased association in a linear model rather than causal relationship between the treatment and the outcome. These assumptions imply that the ANCOVA accurately

[4] This assumption requires that the expected value of the error term, conditional on the independent variables, is zero, $E[\,\varepsilon_1 \mid T, X\,] = 0$. This assumption automatically ensures that the errors are uncorrelated with all predictors in the model, which is a standard assumption of the error term, $Cov(T, \varepsilon_1) = 0$ and $Cov(X, \varepsilon_1) = 0$. Beyond this standard assumption of the error term, zero conditional mean assumption further ensures that the error term does not systematically vary with any function of the regressors, thereby covering both linear and non-linear relationships. While this assumption is stronger than the standard assumption of the error terms, it is important for the unbiasedness of Ordinary Least Squares (OLS) estimators and accurate interpretation of regression coefficients in modern regression analysis in which the regressors are random variables, not fixed values.

captures the association between the treatment and the outcome when all the variables are observed, but it does not inherently capture the unbiased treatment effect under specific causal structures of variables, particularly when the unobserved confounders are considered in causal inference frameworks. For example, the zero conditional mean of the error assumption on multiple linear regression model where ANCOVA is one of them, which states that the expectation of the error term is zero given the independent variables (i.e., $E[\varepsilon_1 \mid T, X] = 0$), could ensure the unbiasedness of the OLS estimate of the regression coefficient for the treatment grouping variable, T, if the model is correctly specified. However, if the model is not correctly specified (e.g., there was a relevant confounding variable, but the variable was excluded from the model), then it leads to the omitted variable bias for estimating the treatment effect (Wooldridge, 2020, pp. 84-85).

2.1.3. Change Score Analysis

In the two-wave panel designs, the change score analysis or the gain score analysis offers an alternative for examining the treatment effect in addition to the ANCOVA. Change score analysis has an advantage over ANCOVA to determine if there was significant change within a specific group from pretest to posttest. It is particularly useful in applied situations where researchers want to assess overall improvement within the specific treatment group, such as evaluating the effectiveness of the specific program without the control group, which are prevalent in the fields of Education and Psychology. [5] ANCOVA is not ideal for this purpose since it focuses on differences between groups rather than within them (Maxwell et al., 2018, p. 499). However, the change score analysis, of course, can be used to compare treatment effects between groups and was applied in the Head Start program in a situation where an untreated control group existed. The

[5] If the available data has a single treatment group, such as a pretest and posttest on the same treatment measure, then the question of whether there has been significant change from pretest to posttest is answered by performing what a matched-pairs or dependent *t*-test of the data is essentially.

change score analysis applied in the Head Start program compares these differences in outcomes from pre-test to post-test between a Head Start group and a control group, thereby estimating the treatment effect. The regression model of the change score analysis for the two-wave data scenario can be defined as follows:

$$Y - X = \beta_0 + \gamma_{\text{Change}} T + \varepsilon_2 \quad (2)$$

where the difference between the posttest score Y and pretest score X is regressed on the dichotomous treatment variable T. The β_0 is the regression intercept and the change score analysis estimator γ_{Change} is the regression coefficient for T, and it is assumed that T and the error term, ε_2, are uncorrelated, $Cov(T, \varepsilon_2) = 0$. Apart from the consideration of with or without random assignment procedure in the research design, statistical OLS estimation of γ_{Change} in equation (2) is the equivalent procedure as an independent *t*-test on change scores between the treated and untreated group.

2.1.4. Comments on Performances of the Two Analytic Methods

Maxwell et al. (2018, pp. 498-502) argued that the two traditional methods test different null hypothesis depending on the presence or absence of random assignment. When participants in a study are randomly assigned to the different groups (i.e., treatment or control groups), each group is statistically equivalent before the treatment so that any observed differences in outcomes between the groups after the treatment can be attributed to the treatment itself, rather than pre-existing differences between the groups. In other words, with random assignments, both analytic methods aim to test the same null hypothesis regarding the equivalent groups in all respects. For example, if we assume, contrary to the fact, that random assignment was performed in the Head Start program example, both methods test the same null hypothesis, there is no difference in the group means of post-test scores between the Head Start and control groups. However, they added

that even under the random assignment, “ANCOVA is almost always preferable because it provides more power and precision” (Maxwell et al., 2018, p. 501; Venter, Maxwell, & Bolig, 2002). From a model comparison perspective, equation (2) can be seen as a constrained form of equation (1), where the regression slope of the postscores on the prescores is fixed at 1, indicating a perfect one-to-one relationship between the prescores and the postscores. Since the relationship between the pre- and post-test scores could not be perfect in real data, the error variance under change score analysis model tends to be larger compared to the ANCOVA model, where the slope can be adjusted to minimize this error. Consequently, in randomized studies, ANCOVA often offers greater statistical power than the change score analysis, making it a more efficient approach for detecting the treatment effects.

In contrast, without random assignments in a quasi-experimental design such as the actual Head Start example, Maxwell et al. (2018) argued that the ANCOVA and change score analysis do not test the same null hypothesis but test different hypotheses and researchers' failure to understand this difference is known as Lord's paradox. Without random assignment, ANCOVA tests the conditional null hypothesis, if the groups were adjusted to be equivalent on the pretest scores, there is no difference in the group means of posttest scores between the Head Start and control groups. To put it more specifically in the form of a question, “For Head Start participants and non-participants having identical pretest scores in September, would a difference in mean posttest scores between groups in June be expected?” On the other hand, Maxwell et al. also suggested that change score analysis could be the preferred method in studies where the specific conditional inquiries addressed by ANCOVA are not relevant, particularly in intact group settings (e.g., “ethnic or cultural groups” in p. 454) where the primary interest is in simple comparison under change scores between groups without considering initial pretest scores. Overall, Maxwell

et al. argued that in nonequivalent control group designs such as Lord's paradox, there is no uniformly superior method, but it depends on the research question.

2.2. Graphical Causal Models on Two-wave Designs

Beyond traditional methods, this dissertation focuses on how unobserved confounders bias treatment effect estimates in two-wave nonequivalent control group designs, comparing ANCOVA and change score analysis. Holland and Rubin (1983) revisited four examples, including Lord's Head Start study, and showed that Head Start was not a randomized causal study because program assignment depended on group membership (for example, disadvantaged status). In that example, pretest scores influenced who entered the Head Start program or the control group, creating preexisting differences before treatment. Pretest scores could also directly or indirectly affect posttest scores through an unobserved mediator which is independent of program participation. Thus, the pretest score could affect both treatment assignment and the posttest outcome. When an additional unobserved variable influences pretest, treatment assignment, and posttest, traditional methods struggle to clarify the assumptions needed for an unbiased treatment estimate. This difficulty highlights the need to bring unobserved confounders into the graphical causal model to properly address the structural biases.

Graphical causal model addresses this challenge by visualizing all relevant variables and causal relations in directed acyclic graph (DAGs). Nodes represent variables such as pretest score, treatment assignment and posttest score, and arrows denote direct causal effects. Confounding bias appears as any back-door path between the treatment and the outcome; conditioning on appropriate covariates or designing the study to close these paths ensures identification of the treatment parameter. Moreover, by isolating the structural component of bias arising from unblocked back-door paths, this approach permits a partial quantification of structural bias under ANCOVA. This

level of clarity makes explicit the causal assumptions needed for valid estimation of the treatment effect in two-wave nonequivalent control group designs.

2.2.1. Graphical Causal Models

Judea Pearl's graphical causal model provides a major framework for causal inference alongside the potential outcomes framework (Pearl, 2009, 2016). The central idea is to represent qualitative causal assumptions using graphs, where variables are depicted as nodes and causal relations are depicted as directed edges (arrows). Let denote T for the treatment/intervention, X for the pretest (baseline outcome), Y for the posttest (follow-up outcome), and A for an participant's characteristic that may affect both treatment assignment and outcomes.

A graph consists of nodes connected by edges. A path is a sequence of edges connecting a set of distinct nodes. When the edges have directions, the path is called a directed path, and the graph is called a directed graph. A Directed Acyclic Graph (DAG) is a directed graph that contains no directed cycles. That is, starting from any node and following the direction of arrows, it is impossible to return to the starting node. Figure 1 contrasts an acyclic directed graph with a cyclic directed graph, highlighting that cycles correspond to feedback structures that violate the acyclicity condition required for a DAG (Pearl et al., 2016). In many applications, DAGs are assumed "causally ordering", meaning a relationship in which one variable causally influences another through a directed path, defined as a sequence of connections where arrows consistently point in a single direction, from cause to effect. When combined with a linearity assumption, a DAG can be regarded as the graphical model of a Structural Causal Model (SCM), in which each endogenous variable is generated by a structural function of its causes with an error term (Pearl, 2016).

Within DAGs, a directed path from T to Y, $T \rightarrow Y$, represents a causal path. However, not all paths connecting T and Y correspond to causal influence. A central concept of noncausal

relationship in DAGs is the backdoor path, which is a noncausal path between *T* and *Y* that begins with an arrow into *T*. Backdoor paths reflect spurious associations induced by unobserved confounders or observed colliders, and they are the source of bias when estimating treatment effects from observational data (Morgan & Winship, 2014; Pearl, 2009, 2016). Figure 2 depicts the typical confounding structure in which an unobserved variable *A* is confounding the relationship between *T* and *Y*, yielding the backdoor path $T \leftarrow A \rightarrow Y$. If the backdoor paths are not "blocked" and remain open, those can induce systematic structural bias in the estimated association between *T* and *Y* (Cunningham, 2021; Pearl, 2009, 2016).

Graphical causal modeling clarifies how researchers can reduce or eliminate such bias through conditioning (also referred to as controlling or adjusting). Conditioning can be implemented through several common design and analysis strategies, including regression adjustment, matching, stratification, or weighting; in linear settings, "controlling for" a variable in regression corresponds to conditioning on that variable in the graph (Cunningham, 2021). The backdoor criterion formalizes this idea: a set of covariates is sufficient for identifying the causal effect of *T* on *Y* if conditioning on that set blocks all backdoor paths from *T* to *Y* without opening new noncausal paths (Pearl, 2009, 2016; Pearl et al., 2016). In the two-wave outcome setting that motivates this dissertation, a typical analytic choice is to condition on the pretest *X* (as in ANCOVA). Graphically, this corresponds to blocking any backdoor paths that run through *X* as a non-collider, while leaving open any backdoor paths that run through unmeasured confounders such as *A*.

Another fundamental concept in DAGs is a collider. A collider is a variable that lies on a path with two arrows pointing into it. Colliders behave differently from confounders: whereas conditioning on a confounder tends to block backdoor paths, conditioning on a collider can open

a previously closed path and induce spurious association between its causes (Cunningham, 2021; Pearl, 2009, 2016). In evaluation research with two-wave outcomes, a practically important collider example arises from selection or attrition. For instance, let R denote a posttest retention (i.e., remaining in the study until Y is measured). If both T and A (observed or unobserved) affect R ($T \rightarrow R \leftarrow A \rightarrow Y$), then R is a collider. Restricting analysis to posttest completers (conditioning on $R = 1$) can open a noncausal path between T and Y, then it can induce bias in the treatment effect (Cunningham, 2021; Morgan & Winship, 2014).

Taken together, DAGs offer a compact language for stating causal assumptions and for diagnosing when common analytic decisions block or open noncausal paths. This graphical perspective is particularly useful for clarifying why certain estimators from regression-based models (e.g., ANCOVA conditioning on X) may still be biased in the presence of unmeasured confounding or selection mechanisms. Compared with purely algebraic derivations using the OLS, DAGs can provide a more intuitive route to identifying the sources of bias by explicitly displaying the open backdoor paths and collider structures implied by the assumed data-generating process (Pearl, 2013; Pearl, 2009, 2016). For this reason, the subsequent chapters use DAG-based reasoning (backdoor criterion and path-tracing logic) to complement regression-based derivations when comparing ANCOVA and change score estimators under different confounding scenarios.

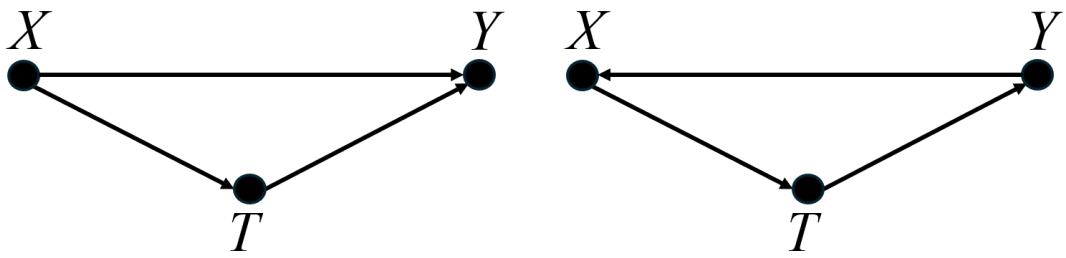

Figure 1. A directed acyclic graph on the left and a cyclic graph on the right.

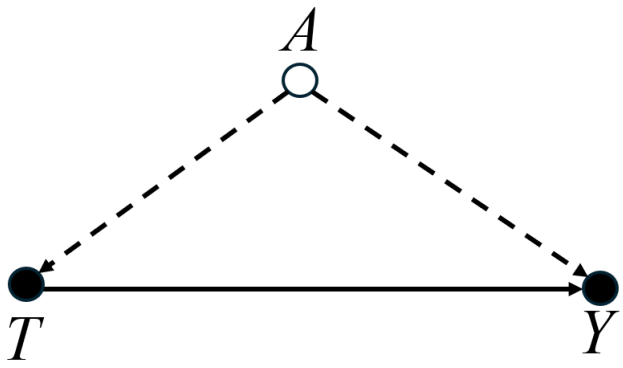

Figure 2. A confounding structure: an unobserved confounder *A* opens a backdoor path between treatment *T* and outcome *Y* when *A* is not conditioned on.

2.2.2. Applications for Two-wave Nonequivalent Control Group Designs

This chapter reviews the previous research that applied the graphical causal model to the two-wave nonequivalent control group design, particularly in the context of Lord's paradox comparing the changes score analysis and the ANCOVA.

Although Pearl (2016) presented the application of DAG to identify the effect of sex on weight in original example of Lord's paradox, there are limitations in deriving general assumptions for obtaining unbiased treatment effects in a two-wave nonequivalent control group design. First, sex was not manipulable treatment/intervention variable as mentioned in Section 2.1. Second, it did not assume the potential unobserved confounder that could confound the relationship between the treatment/intervention variable and the outcome. Lastly, although it showed the application of path-tracing rules to identify the effect of sex, the theory of backdoor criterion, a key tool for identifying systematic bias (i.e., confounder or collider), was not addressed enough. Therefore, Pearl's (2016) study has limitations in clarifying the assumptions needed to obtain an unbiased treatment effect where an unobserved confounder exists.

Kim and Steiner (2021) suggest that, despite requiring a strong common-trend (time-invariant confounding) assumption, change score analysis can be advantageous relative to

ANCOVA in two-wave pretest–posttest settings precisely because it is not exposed to the bias amplification and collider-bias risks created by conditioning on the pretest. At the same time, they emphasize that this advantage comes at the cost of a strong identifying requirement, often framed as a common-trend (i.e., time-invariant confounding) assumption. Using a series of DAGs, they adapt Pearl's (2013) path-tracing logic of DAGs to the change score setting and show that, because the change score estimator is a simple regression coefficient, enumerating open noncausal paths and summing their path products reproduces the exact OLS bias expression. They also present an algebraic expression for an ANCOVA estimator in which the posttest outcome is regressed on treatment while controlling for the pretest, and they interpret ANCOVA's bias primarily through the lens of bias amplification: any remaining confounding bias can be scaled up by an amplification-type denominator term when the pretest both causally affects treatment selection and fails to fully remove confounding.

Despite these contributions, there remain unresolved pieces in their DAG-based use of path-tracing for identifying the structural bias of treatment estimator in ANCOVA. First, while the paper walks the reader through an explicit DAG-based use of path-tracing for change score analysis, the corresponding ANCOVA estimator expression is stated as a result and the paper directs readers to an appendix for derivations, rather than presenting a parallel step by step tracing based graphical approach in the main text. Second, because the ANCOVA estimator is a partial regression coefficient, its population-level estimator can be written as a regression of the outcome on a residualized version of treatment, which introduces a variance scaling term induced by partialling out the pretest. Kim and Steiner (2021) discuss amplification through the denominator of the ANCOVA estimator bias term, but the connection between this denominator and the residualization implied by partial regression is not made explicit as a practical workflow that

separates what can be read directly from the graph and what requires an algebraic partialling out step. Third, although the paper positions graphs as a powerful complement to algebraic formulations for understanding bias mechanisms, it does not frame as a focal question whether the same DAG-based use of path-tracing reproduces the OLS estimation exactly for both the change score and the ANCOVA estimators, or whether the correspondence differs across estimators due to residualization induced scaling. These points leave room for a more estimator-focused synthesis that makes the following elements explicit: how far the DAG-based use of path-tracing can be carried over from simple regression such as change score analysis to multiple regression such as ANCOVA; which components of structural bias are transparently attributable to open backdoor paths; and which components arise from residualization and therefore need to be handled via the population OLS estimation.

Lüdtke and Robitzsch (2025) discussed the assumptions required for obtaining unbiased treatment effect estimates under ANCOVA and change score analyses in two-wave outcome data. Their argument is framed in a causal data-generating perspective in which an outcome is measured at two occasions, with the first wave serving as the pretest and the second wave serving as the posttest. Treatment assignment (T) is allowed to be confounded by an unobserved variable (U) that affects both outcome waves as well as treatment assignment, reflecting the possibility that unmeasured characteristics (e.g., ability or motivation) drive both selection into treatment and outcomes over time. In addition, the baseline outcome may predict treatment assignment (capturing selection processes in nonequivalent group designs) and may also predict the posttest outcome (capturing outcome stability over time). Within this general setup, the estimator of interest is the causal effect of T on the posttest outcome.

To obtain interpretable expressions for the implied estimators, Lüdtke and Robitzsch (2025) adopt standard simplifying assumptions commonly used in linear SCM discussions: effects are linear and additive (i.e., no interaction terms), disturbances are mean-zero and mutually uncorrelated and are independent of the predictors, and variables are treated as measured without error (with several variables standardized for convenience). Under these assumptions, they derive the population regression coefficient on *T* implied by each analytic method when the unobserved confounder *U* is omitted from the fitted model, and they characterize the resulting systematic (structural) bias.

For ANCOVA, the analytic model regresses the posttest outcome on treatment while adjusting for the pretest. Lüdtke and Robitzsch (2025) show that when *U* affects both treatment assignment and the posttest outcome, the ANCOVA estimator generally differs from the causal effect of *T* on the posttest outcome because conditioning on the pretest alone does not, in general, eliminate confounding due to *U*. Adjusting for the pretest can reduce bias to the extent that the pretest is informative about *U* (i.e., the pretest partially proxies the unobserved confounder), but it does not guarantee unbiasedness. In this sense, unbiased ANCOVA estimation requires that there be no remaining unmeasured confounding between *T* and the posttest outcome (e.g., *U* does not affect treatment assignment or does not affect the posttest outcome, or *U* is measured and included among the covariates).

They also analyze the change-score analysis, in which the treatment effect is estimated using the difference between posttest and pretest outcomes as the outcome variable. Their key point is that differencing does not automatically remove bias induced by an unobserved confounder; rather, unbiasedness requires additional structure. In particular, they emphasize that two substantive conditions are needed for the change score estimator to be unbiased in the presence of

an unobserved confounder *U*: first, a time-invariance condition on confounding such that the influence of *U* on the outcome process is stable across the two measurement occasions (so that its contribution cancels in the change score), and second, an absence of "dynamic" selection in which prior outcomes directly influence treatment assignment (i.e., treatment assignment is not driven by the baseline outcome). They connect the first condition to the common-trend assumption (e.g., Kim and Steiner, 2021; Lechner, 2010) and the second condition to the assumption of no dynamic causal relationships between prior outcomes and treatment assignment (e.g., Imai and Kim, 2019).

Although Lüdtke and Robitzsch (2025) present the causal structure underlying their data-generating perspective, their bias analysis is developed primarily through ordinary least squares (OLS) derivations within a traditional structural equation modeling framework, rather than through graphical causal modeling tools such as the backdoor criterion or path-tracing rules. While the OLS estimation provides a direct algebraic route to expressing systematic bias, it can become cumbersome as the confounding structure grows more complex (e.g., when multiple confounders or additional causal paths are introduced). A complementary DAG-based graphical perspective can make the sources of structural bias more explicit by indicating which noncausal paths remain open under a given analytic model (e.g., conditioning only on the pretest) and how those open paths generate bias in the treatment effect estimator. This graphical interpretation can therefore provide a more intuitive framework for diagnosing and communicating when and why ANCOVA or change score estimators become biased, particularly in more elaborate two-wave design settings.

2.2.3. Pretest as a Mixture of Stock and Flow Components Perspective

In two-wave designs, the decision to model a causal effect from the pretest to the posttest must be theoretically justified rather than based solely on predictive associations in regression analyses. This distinction matters because prediction and causation are conceptually different: a

strong predictive association between the pretest and posttest does not, by itself, warrant a causal edge $X \rightarrow Y$. Instead, positing a direct causal arrow requires an account of what component of the pretest plausibly persists over time or what process is newly generated at the pretest wave and then affects the later outcome. Recent two wave causal analyses already treat a direct causal arrow from the pretest X to the posttest Yas plausible (Kim & Steiner, 2021; Lüdtke & Robitzsch, 2025). However, these models often include the $X \rightarrow Y$arrow without a detailed account of what kind of causation is being assumed, and they do not explicitly explain why such a direct effect could remain in the presence of an unobserved, time invariant confounder such as latent ability A. Kim and Steiner's discussion also illustrates that the plausibility of $X \rightarrow Y$depends on how the pretest is conceptualized and on what between wave processes are deemed credible. In their initial diagrams that introduce latent ability A to represent stable individual differences, the direct $X \rightarrow Y$ edge is not always needed to explain pretest posttest association, which aligns with an interpretation in which the pretest mainly reflects stable, stock like differences. They later describe cases in which the pretest score becomes known after the pretest, and this information can trigger processes such as increased student motivation or parental engagement (for example, arranging private tutoring), which makes a direct pretest to posttest relation more plausible even when a stable ability confounder is present. These examples align naturally with a flow interpretation in Allison's (1990) sense, because the key driver is a response that is newly generated at the time point (or immediately after it) and then carries forward to influence the later outcome, even though the mechanism is not represented explicitly with a mediator in the graph.

Allison (1990) offers a useful conceptual entry point by framing the pretest to posttest relation using the distinction between stocks and flows, a terminology that originates in economics. Stocks are quantities that inherently persist over time (for example, organizational size and body

weight), whereas flows are quantities that must be continuously generated at each time point (for example, many measures of behavior and attitude). Although Allison argues that a direct pretest to posttest causal path is most straightforward for stocks, he also notes that flows can plausibly exert a direct influence on later outcomes when they are subject to reinforcement or habit formation. This stock and flow framework is valuable precisely because it pushes the discussion beyond regression prediction and toward causal interpretation: it forces the analyst to articulate why a pretest measure should be expected to exert a direct influence on later outcomes in the data generating process. From this perspective, Allison's stock versus flow distinction can be used as a dichotomous heuristic for motivating a direct pretest to posttest path. The pretest is most plausibly treated as having a direct effect when it is conceptualized as a stock that persists over time, or as a flow that is newly generated at the wave and then carries over through reinforcement or habit formation. When this rationale is weak, it may be more appropriate, in SEM terms, to treat the pretest primarily as a reflective indicator of an underlying latent attribute rather than as a formative cause of the posttest. In real world practice, however, pretest scores in education or psychology seldom reflect a single, pure stock or a single, pure flow variable. Whether the pretest measures cognitive abilities or behaviors and attitudes, it often blends stock-like components (for example, stable ability or long-standing test taking proficiency) with flow-like components (for example, current interest, engagement, or situational motivation). Because the exact proportions of these components are not directly observable, it is often unrealistic to force a binary classification. I therefore adopt the pretest as a mixture perspective: the pretest contains an unknown blend of stock-like and flow-like components.

The mixture perspective can be further refined by adding the idea of accumulation. The pretest score reflects influences that have accumulated prior to the pretest, but the stock and flow

components differ in how their effects propagate across waves. The stock component represents relatively stable attributes that have accumulated before the pretest and therefore carry over naturally. The flow component, by contrast, is newly generated at the time point (or in response to time specific information such as learning the pretest score) and may then accumulate between waves through reinforcement, habit formation, and other processes that unfold after the pretest. This accumulation idea is consistent with longitudinal SEM work on what Usami, Murayama, and Hamaker (2019) term "accumulating factors," where common factors are not separated from lagged relations and therefore have both direct and indirect effects that accumulate through the lagged structure. Although such accumulation processes generally require more than two waves to be empirically disentangled, the accumulation concept provides a useful theoretical supplement for motivating why an $X \rightarrow Y$ relation may remain plausible in a two-wave design even after introducing a time invariant ability confounder.

Under the pretest as a mixture perspective, the direct edge $X \rightarrow Y$ can plausibly persist even after adding an unobserved confounder for latent ability A, assumed to be time invariant over the pre to post period, to the DAG. A direct edge $X \rightarrow Y$ may legitimately remain because the pretest can contain flow like components that are not fully explained by A and can causally influence the posttest, including wave specific responses triggered by the pretest information itself (for example, changes in motivation after learning the pretest score). At the same time, introducing a mediator M can make this intuition more explicit by representing a concrete between wave mechanism through which these newly generated flow components accumulate and ultimately affect the posttest. In this representation, the path $X \rightarrow M \rightarrow Y$ captures a plausible process that unfolds between waves (for example, pretest triggered motivational factor such as self-efficacy leading to sustained engagement that builds over time and improves posttest performance), while retaining

$X \rightarrow Y$ allows for remaining carryover that is not explicitly modeled or is difficult to isolate with only two waves. In this sense, Kim and Steiner's examples can be interpreted as consistent with a mixture-oriented account, in which stable trait influences are captured by A while additional between-wave processes initiated around the pretest occasion can motivate retaining a direct $X \rightarrow Y$ relation, even when those processes are not represented with an explicit mediator. The pretest as a mixture perspective makes this logic explicit, provides a clearer rationale for when a direct pretest to posttest effect may remain in the presence of an ability confounder, and clarifies when mediation mechanisms between waves should be represented directly in the DAG.

Based on these theoretical backgrounds and identified gaps, this dissertation aims to explore how graphical causal models can be employed to identify structural biases in causal treatment effects within two wave nonequivalent control group designs. Traditional analytic methods, such as ANCOVA and change score analysis, often fail to detect biases caused by unobserved confounders, because these biases are typically subsumed in the error term when the analytic model diverges from the true data generating model in the population. In contrast, graphical causal models provide more intuitive and effective tools for identifying systematic biases. To fill the research gaps identified in previous studies, this dissertation offers a more accessible and intuitive framework for identifying structural biases. Specifically, it demonstrates how graphical causal tools, such as the backdoor criterion and path-tracing rules within directed acyclic graphs (DAGs), can simplify the identification and interpretation of structural bias for both ANCOVA and change score analysis, especially in various two wave data generating models. Relying solely on the OLS estimation becomes cumbersome and less transparent under such conditions, as it involves detailed algebraic derivations that may not be readily understood by substantive researchers. By showing when and why certain established assumptions (e.g., the

common trend assumption in change score analysis and no unmeasured covariates in ANCOVA) are required to obtain the unbiased treatment effect, and how each analytic method can open or close potential backdoor paths, this dissertation addresses key limitations of earlier work. Moreover, adopting the pretest as a mixture perspective provides a rationale for the DAGs analyzed in the scenarios that follow in Methodology: it explains why a direct pretest effect may persist in the presence of an ability confounder, and it motivates the inclusion of a mediator when theory (e.g., self-efficacy) suggests a mechanism between waves.

3. Methodology

This chapter addresses the methods to identify the structural biases in two-wave, nonequivalent control group designs through the application of the graphical causal model. Section 3.1 focuses on how to identify structural biases, utilizing tools such as the backdoor criterion and path-tracing rules to visually detect and address biases arising from both unobserved confounders and the analytic method itself. Section 3.2 introduces various scenarios represented by directed acyclic graphs (DAGs), which depict different causal structures within two-wave data. These scenarios will be used to explore how biases manifest under different conditions and what assumptions are necessary to obtain the unbiased treatment effect estimates.

3.1. Identifying Structural Bias

This dissertation uses the term *structural bias* to denote the population-level difference between the causal treatment effect defined by a correctly specified structural causal model (i.e., γ in the data-generating model) and the treatment effect estimated by the chosen analytic model (i.e., γ_{ANCOVA}, the population-level ANCOVA estimator, or γ_{Change}, the population-level change score estimator). I then present two distinct approaches to identifying this structural bias: the graphical approach, which uses the backdoor criterion to list open noncausal backdoor paths and path-tracing rules to express the induced noncausal association as products of structural coefficients, and the OLS estimation at the population level, which derives the corresponding bias using regression formulas, including the Frisch–Waugh–Lovell (FWL) theorem for ANCOVA. In line with the definition of structural bias, this dissertation focuses solely on the direct effect of T on Y as the treatment effect. Any additional causal effect of T on Y that operates through mediators

in the data-generating model but is omitted from the analytic model is therefore treated as structural bias.

3.1.1. Backdoor Criterion

Understanding the backdoor criterion and its role in identifying bias requires an exposition of the logical progression from blocked paths to *d*-separation (the *d* stands for "directional") and then to conditional independence in graphical causal model. Let X and Y be random variables representing the treatment and outcome, respectively, and let Z be a set of random variables that conditioned on. According to Pearl et al. (2016, p. 46), *"Two nodes X and Y are d-separated if every path between them (should any exist) is blocked."* This means that when all paths connecting X and Y are blocked, they are *d*-separated in the DAG. This introduces *d*-separation as a graphical criterion indicating that X and Y are *d*-separated when all possible paths connecting them in the DAG are blocked. Pearl et al. further provided the general definition of *d*-separation: *"If Z blocks every path between two nodes X and Y, then X and Y are d-separated, conditional on Z, and thus are independent conditional on Z"*. This definition connects the graphical concept of blocked paths with the statistical concept of conditional independence, implying that if a set of variables Z blocks all paths between X and Y, then conditioning on Z makes X and Y statistically independent, which means X and Y are *d*-separated. In contrast to *d*-separation, two nodes X and Y are *d*-connected if there exists at least one unblocked path between them in the DAG after conditioning on the set of variables Z. This means that not all paths between X and Y are blocked by Z, allowing for potential dependencies between X and Y given Z. When X and Y are *d*-connected, associations between X and Y can transmit through the unblocked paths as they are not conditionally independent given Z.

This *d*-separation, implying that X and Y are conditionally independent given Z, can be expressed mathematically as:

$$P(X, Y \mid Z) = P(X \mid Z) \cdot P(Y \mid Z).$$

This equation means that once the researcher knows Z, the conditional joint probability distribution of X and Y conditional on Z factors into the product of their individual conditional distributions. In other words, conditioning on Z removes any association between X and Y that is transmitted from the variables in Z, aligning with the concept of blocking paths in the DAG. Additionally, this conditional independence implies that:

$$P(X \mid Y, Z) = P(X \mid Z)$$

or

$$P(Y \mid X, Z) = P(Y \mid Z).$$

These mathematical expressions mean that once Z is known, knowing Y provides no additional information about X, and vice versa. In other words, conditioning on Z makes X and Y independent of each other. Thus, the mathematical framework supports the concept of blocking by demonstrating how conditioning on certain variables affects probability distribution and independent relationships among the variables in the DAG.

The concept of *"blocking"* can be implemented through several conditioning methods, depending on the study's context and the feasibility of observing confounders. One of the most well-known methods for achieving blocking is through controlling for covariates in regression analysis, which conditions on these covariates by including them in the model to isolate the causal relationship of interest. By adjusting for relevant covariates, regression analysis effectively blocks non-causal paths between the treatment and outcome variables. Similarly, matching and stratification (also known as subclassification) are conditioning methods that can be used when random assignment is not feasible. Matching involves pairing units with similar values on the confounders to ensure balanced comparison groups, effectively conditioning on those variables by

equalizing their influence across groups. Stratification divides the sample into homogeneous subgroups (strata) based on the confounders and compares outcomes within each stratum, thereby conditioning on those confounders by holding them constant within each group.

In the graphical causal model, by definition, a path between X and Y is considered blocked by Z if it satisfies certain conditions based on the structures within the DAG. Specifically, a chain structure occurs when the path follows a sequence like $A \rightarrow B \rightarrow C$ and a fork occurs when a node has two outgoing edges, such as $A \leftarrow B \rightarrow C$. In both chains and forks, conditioning on the middle node B blocks the path. By including B in the conditioning set Z, it can prevent any association from being transmitted between A and C through B. This leads to the conditional independence of A and C given B, expressed as $A \perp\!\!\!\perp C \mid B$. The fork structure is fundamental when dealing with confounders[6] in which variables influence both the treatment X and the outcome Y. For example, in the structure $X \leftarrow Z \rightarrow Y$, Z is a confounder creating a backdoor path that can bias the estimation of the causal effect of X on Y. By conditioning on Z, it blocks this backdoor path, and it leads to conditional independence between X and Y given Z (i.e., $X \perp\!\!\!\perp Y \mid Z$) ensuring that any spurious association between X and Y due to Z is removed, allowing us to isolate the causal effect of X on Y.

A collider, on the other hand, is a structure where two paths converge on a single node, forming $A \rightarrow B \leftarrow C$. In this case, node B is a collider because it is influenced by both A and C. Not conditioning on collider B or any of its descendants keeps the path between A and C naturally blocked (i.e., A and C are d-separated). By not including collider B in the conditioning set Z, the

[6] In this context, the term confounder refers to a variable that affects both the treatment and the outcome, introducing bias that needs to be addressed to obtain an unbiased estimate of the causal effect. Confounders may be either observed or unobserved. When a confounder is unobserved, it poses additional challenges because it cannot be directly controlled in the analysis, whereas observed confounders can be adjusted for as covariates in a regression model.

collider prevents any association between *X* and *Y*, making them marginally independent without given *Z*, that is expressed following

$$P(X, Y) = P(X) \cdot P(Y).$$

In contrast, conditioning on collider *B* or its descendants unblocks the path, creating a spurious association between *A* and *C* that did not exist before (i.e., *A* and *C* are *d*-connected). Therefore, by conditioning on the appropriate set *Z* that blocks all paths between *X* and *Y*, we ensure that *X* and *Y* are conditionally independent given *Z*.

Having established how blocked paths and *d*-separation lead to conditional independence between variables in a DAG, the backdoor criterion is introduced, which leverages these concepts to facilitate causal inference. The backdoor criterion provides a systematic way to determine a set of variables *Z* that, when conditioned upon, blocks all backdoor paths between the treatment *X* and the outcome *Y*, thereby allowing for an unbiased estimation of the causal effect of *X* on *Y*. Pearl et al. (2016, p. 61) originally defined the backdoor criterion in context of identifying treatment/intervention effects: *"Given an ordered pair of variables (X, Y) in a directed acyclic graph G, a set of variables Z satisfies once every possible backdoor path has been blocked, it is considered that the research design has met the backdoor criterion relative to (X, Y) if no node in Z is a descendant of X, and Z blocks every path between X and Y that contains an arrow into X".* The *X* and *Y* are defined as the treatment and the outcome respectively, and *Z* is the set of confounders that need to be controlled to accurately estimate the causal effect of X on Y. Thus, this definition means that to meet the backdoor criterion, the set *Z* should be carefully chosen as a set of confounders (or "conditioning set") that can account for any non-causal association between treatment *X* and outcome *Y*. None of the variables in *Z* should be influenced by treatment *X*, and

conditioning on *Z* must block all non-causal paths between the treatment *X* and the outcome *Y* that could otherwise introduce bias.

Pearl (2016, pp. 61–62) proposed that *Z* must satisfy the following three conditions to meet the backdoor criterion**.** First, all spurious paths between treatment *X* and outcome *Y* must be blocked. This means that *Z* should be chosen to block any non-causal (backdoor) path connecting *X* to *Y* through confounders. These paths, which begin with an arrow pointing into *X***,** may create spurious associations between *X* and *Y* by carrying the influence of other factors. By conditioning on the appropriate variables in Z, these paths can be blocked, ensuring that only the causal effect of *X* on *Y* is estimated. Second, all directed paths from treatment *X* to outcome *Y* must remain unperturbed. This condition ensures that the direct causal path from *X* to *Y* stays intact. Conditioning on variables along this causal path or on descendants of *X* might inadvertently block or distort the causal influence from *X* to *Y*. Therefore, *Z* should not include any variables that would interfere with this directed path, allowing the causal relationship to be accurately estimated. Third, no new spurious paths should be created. This condition requires that conditioning on Z does not introduce new associations between treatment *X* and outcome *Y*. Specifically, conditioning on colliders (nodes where two arrows converge head-to-head, such as $X \rightarrow M \leftarrow Y$) can open previously blocked paths, creating a new association between *X* and *Y* that does not represent a causal relationship. To satisfy this condition, *Z* should exclude colliders and their descendants to prevent these unintended paths from introducing bias.

However, when certain confounders that should be included in *Z* are unobserved or cannot be measured, the backdoor criterion still plays a role in identifying potential biases in estimating the causal effect of *X* on *Y*. By mapping out all possible backdoor paths in the DAG, researchers can pinpoint which unmeasured confounders may introduce bias through these paths.

Understanding the structure of the DAG allows us to recognize that, without controlling these unobserved confounders, the backdoor paths remain unblocked, leading to a biased estimation of the treatment effect. Therefore, even when researchers cannot observe certain confounders in *Z*, the backdoor criterion helps identifying the biases for the causal effect of *X* on *Y*.

3.1.2. Path-Tracing Rules in DAGs

Sewall Wright (1921) introduced path analysis with the core idea of path-tracing rules. This method involves constructing path diagrams in which variables are depicted as nodes and directed arrows represent hypothesized causal relations. Under the standard linear and recursive formulation, the associated path coefficients can be interpreted as standardized regression coefficients that index the magnitude and direction of direct effects between variables. Wright's key insight was that the overall association between two variables can be decomposed into a sum of path-based components by tracing all admissible paths between the two variables and summing the products of the coefficients along those paths.

The notion of an admissible path, however, differs across traditional path analysis in SEM and causal analysis using directed acyclic graphs, because the rules for when a path transmits association depend on the graph structure and on whether variables are conditioned on. A widely used SEM oriented statement of tracing rules is given by Kenny (1979, pp. 37 to 38) for recursive models[7]. In that formulation, the correlation between two variables can be obtained by summing the products of path coefficients over all admissible tracings that satisfy two conditions: (a) the same variable is not entered more than once in any single tracing, and (b) a variable is not entered

[7] Unidirectional, acyclic relationships among variables characterize a recursive model in SEM. In such models, causal effects flow in one direction from exogenous variables to endogenous variables without any reciprocal causation between endogenous variables. This means that each endogenous variable is caused only by variables that precede it in the causal order, and the disturbances associated with the endogenous variables are assumed to be uncorrelated (Kenny, 2023, p. 109).

through an arrowhead and left through an arrowhead. Condition (a) is consistent with the restriction against repeating nodes in DAG based tracing. Condition (b), in effect, excludes head-to-head meetings and therefore treats collider structures as blocking association in the tracing calculation. In causal DAGs, by contrast, whether a collider blocks association depends on conditioning. Without conditioning, a collider blocks the path, but conditioning on the collider, or on one of its descendants, can open the path and induce association between its causes. Thus, a path such as $X \rightarrow Y \leftarrow Z$ is blocked marginally but can become open if Y is conditioned on, and the relevant set of paths for association and bias analysis is therefore governed by *d*-separation and d connection rules relative to a conditioning set.

Pearl (2013) reintroduces Wright's tracing logic within a modern causal graph framework by using linear structural causal models as a microscope for understanding how causal assumptions imply statistical associations. In this setting, the tracing logic is used to compute covariances, or correlations under standardization, by expressing a covariance as a sum of products of structural coefficients along *d*-connected paths, together with any required error covariances when disturbances are correlated. Pearl also emphasizes that when variables are not all standardized, the same tracing logic can be applied to covariances by accounting for the variance of the appropriate root variable along a traced route. In this dissertation, the focus is on covariances rather than correlations, and the regressors are standardized for convenience while the posttest outcome is not necessarily standardized. Under this scaling choice, many root variance multipliers equal one for the regressors, which simplifies the tracing calculations while preserving interpretability on the outcome scale. Importantly, Pearl also shows how tracing can be used beyond unconditional covariances by combining traced covariance expressions with standardized partial regression coefficient formula to obtain the exact quantity of partial regression coefficients. The resulting

workflow is general and correct, but it typically proceeds by first deriving a set of pairwise covariances implied by the graph and then substituting them into a partial regression formula. For applications such as ANCOVA, where the estimator itself is defined after conditioning on the pretest, this substitution approach can leave the residualization step implicit, because the rescaling induced by residualizing on the pretest is embedded in the algebra of the partial regression coefficient formula rather than being highlighted as a separate, interpretable component.

Kim and Steiner (2021) provide a more directly estimator-oriented use of Pearl (2013) tracing logic in two-wave designs to identify bias for change score estimator. Their approach makes the bias structure interpretable by enumerating open backdoor paths that link treatment and the change score outcome and by summing the products of coefficients along those paths under linear assumptions. At the same time, their analysis is centered on change score estimator, where the target coefficient is a simple regression coefficient and where the graphical and algebraic expressions often coincide under standardization. Their presentation does not provide an explicit extension of the same tracing logic to ANCOVA, where the estimator is a partial regression coefficient and where an additional residualization induced scaling component becomes central for quantifying the exact magnitude of structural bias. This distinction matters for estimator interpretation. When the estimator is a simple regression coefficient, as in change score analysis, the tracing-based decomposition is especially transparent because the coefficient is directly tied to a single covariance normalized by the treatment variance. Kim and Steiner exploit this simplicity by enumerating the open noncausal paths linking treatment and the change score and quantifying each transmitted association as a product of the corresponding structural coefficients. Under this structure, the sum of the noncausal path products maps directly into the expected change score estimator, yielding an immediate unbiasedness condition such as the common trend assumption,

equality of the confounder's effects on the pretest and posttest, which produces the offsetting mechanism emphasized in their application of DAGs.

In contrast, ANCOVA targets a partial regression coefficient after conditioning on the pretest, so the same path product logic does not, by itself, determine the exact bias magnitude. The graph remains sufficient to determine which noncausal paths remain open after conditioning, and therefore to recover the qualitative unbiasedness conditions that correspond to eliminating all remaining open backdoor structures. However, the magnitude of any remaining confounding component depends additionally on how conditioning changes the scale of the estimator through residualization, because the relevant normalization involves residualized variances that are implicit in the partial regression identity. For this reason, the next section complements the tracing based graphical decomposition with the explicit population OLS estimation using the Frisch-Waugh-Lovell theorem, so that the residualization step and its variance scaling role are made explicit and can be cleanly paired with the core confounding component isolated by the graph.

In this dissertation, the combination of the backdoor criterion for identifying the open backdoor paths and path-tracing rules to quantify the bias of corresponding backdoor paths will be referred to as the graphical causal model approach, or simply the graphical approach.

3.1.3. OLS Estimation of Treatment Effect

To identify structural bias, particularly bias arising from unobserved confounders, Ordinary Least Squares (OLS) estimation is employed at the population-level within a structural causal model. Although the OLS estimation is commonly used on sample data to estimate population parameters, focusing on the population regression function helps understand how omitted variables can bias estimates without considering sampling error. By examining the regression function and the associated covariance structure in the population, any correlation

between unobserved confounders and the regressors become clear, revealing the structural bias that causes the treatment effect estimator to deviate from the true causal effect and indicating which causal assumptions must be satisfied to remove this bias. This approach follows Goldberger (1962), who expressed the outcome as a best linear projection onto the observed covariates using the variance-covariance structure. Lüdtke and Robitzsch (2025) applied a similar population-level linear projection framework to identify structural bias in the treatment effect estimator. Working at the population level removes sampling error, thus revealing potential correlations between unobserved confounders and the regressors. For simplicity, this population-level OLS estimation is referred to as the 'OLS' throughout the dissertation. Throughout this dissertation, ANCOVA and change score analysis are discussed at the population level. Accordingly, the corresponding treatment effect estimators are treated as population-level quantities.

The central assumption in this population OLS is the zero conditional mean assumption that the expected value of the error term ε, given the set of covariates X, is zero (i.e., $E[\varepsilon|X] = 0$) mentioned earlier in 2.1.2., from which Cov (ε, X) = 0 and $E[\varepsilon] = 0$ derived. Then these two distinct assumptions are used in derivation of the structural biases under ANCOVA. The first assumption under the zero conditional mean assumption that the error terms are uncorrelated with the set of covariates, i.e., Cov (ε, X) = 0, is fundamental for the unbiasedness of the OLS estimators (Wooldridge, 2020, pp. 84-85).

When considering each bias, it is important to distinguish the "true" data-generating model and the analytic model. The data-generating model represents the true causal processes that generate the observed data. It embodies all the relevant variables and their causal relationships, accurately reflecting how interventions on one variable would cause others. In graphical causal model, the true data-generating model is typically represented by the structural causal model (SCM)

often visualized using Directed Acyclic Graphs (DAGs). The analytic model is the statistical model specified by the researcher, which may differ from the data-generating model due to omitted variables or incorrect assumptions. When there is a data-generating model which is true in the population, and the zero conditional mean assumption holds in this model, then there is no bias for estimating parameters which are regression coefficients. When the analytic model is misspecified and differs from the true data-generating model, the zero conditional mean assumption does not hold because the error term is correlated with the independent variables. If I nonetheless apply the OLS estimation to this misspecified model while assuming the zero conditional mean assumption holds for that model, then bias will emerge. Therefore, to accurately estimate the treatment coefficient, I need to identify the structural bias in each analytic model based on the true data-generating model.

An illustrative example of omitted variable bias arises in the context of estimating the effect of education on income. Suppose the true data-generating model is

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon.$$

where Y is the individual's income, X_1 is years of education, X_2 is innate ability which causes both years of education X_1 and income Y, and ϵ is the error term with the zero conditional mean assumption, $E[\epsilon|X_1, X_2] = 0$. In this true model, both education X_1 and innate ability X_2 are causes of income. However, suppose the researchers omitted the innate ability X_2 from their analytic model because it was unobserved

$$Y = \alpha_0 + \alpha_1 X_1 + u.$$

where the new error term u including both the original error term and the omitted variable's effect, $u = \beta_2 X_2 + \epsilon$. Therefore, the zero conditional mean assumption for the misspecified analytic model, $E[u|X_1] = 0$, is violated because $E[u \mid X_1] = \beta_2 E[X_2 \mid X_1] \neq 0$. The

derivation of the bias for the OLS estimator $\widehat{\alpha_1}$ is needed based on the true data-generating model. The OLS estimator $\widehat{\alpha_1}$ in the underspecified analytic model is

$$\widehat{\alpha_1} = \frac{\mathrm{Cov}(X_1, Y)}{\mathrm{Var}(X_1)},$$

and I can substitute Y into the estimator from the data-generating model,

$$\widehat{\alpha_1} = \frac{\mathrm{Cov}(X_1, \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon)}{\mathrm{Var}(X_1)}.$$

Since the constant do not covary with variables, $\mathrm{Cov}(X_1, \beta_0) = 0$, and

$$\mathrm{Cov}(X_1, Y) = \beta_1 \mathrm{Cov}(X_1, X_1) + \beta_2 \mathrm{Cov}(X_1, X_2) + \mathrm{Cov}(X_1, \epsilon).$$

Since $\mathrm{Cov}(X_1, X_1) = \mathrm{Var}(X_1)$, and $\mathrm{Cov}(X_1, \epsilon) = 0$ from the assumption $E[\,\epsilon \mid X_1, X_2\,] = 0$,

$$\mathrm{Cov}(X_1, Y) = \beta_1 \mathrm{Var}(X_1) + \beta_2 \mathrm{Cov}(X_1, X_2).$$

Then, if I plug it back into the estimator,

$$\widehat{\alpha_1} = \frac{\beta_1 \mathrm{Var}(X_1) + \beta_2 \mathrm{Cov}(X_1, X_2)}{\mathrm{Var}(X_1)} = \beta_1 + \beta_2 \frac{\mathrm{Cov}(X_1, X_2)}{\mathrm{Var}(X_1)}.$$

Since all the terms are constant, the expected value of the estimator is the same,

$$E[\widehat{\alpha_1}] = \beta_1 + \beta_2 \frac{\mathrm{Cov}(X_1, X_2)}{\mathrm{Var}(X_1)}.$$

Therefore, the omitted variable bias is the difference between the expected value of the OLS estimator $\widehat{\alpha_1}$ and the true coefficient β_1 which is,

$$\text{Bias} = E[\widehat{\alpha_1}] - \beta_1 = \beta_2 \frac{\mathrm{Cov}(X_1, X_2)}{\mathrm{Var}(X_1)}.$$

Thus, this derivation of the bias illustrates how the OLS can reveal the bias caused by an unobserved variable when the analytic model differs from the true data-generating model.

Lüdtke and Robitzsch (2025) utilized this method to identify the bias for the OLS estimator of the treatment variable, demonstrating how unobserved confounders can lead to biased estimates even under the zero conditional mean assumption. By incorporating these unobserved factors into

our analysis within the OLS estimation, I can identify the bias associated with the treatment effect estimation. This is because unobserved confounders may cause both the treatment and the outcome, leading to biased estimates even under the zero conditional mean assumption for observed variables. By explicitly considering this potential bias at the population level, I highlight the importance of accounting for all relevant variables in the regression model to accurately estimate the treatment effect.

3.1.3.1. Frisch–Waugh–Lovell (FWL) theorem for ANCOVA

The population OLS estimation under ANCOVA isolates T's unique association with Yafter controlling for the observed covariate X. This clarifies why the population-level ANCOVA estimator for the treatment coefficient can involve additional "residual-variance" components in structural-bias expressions that are not directly visible from a tracing-based graphical characterization alone. The "partialling out" or "residualization" interpretation of multiple regression originated with Frisch and Waugh (1933) and has been widely used in econometrics and causal inference textbooks (Angrist & Pischke, 2009, pp. 35–36; Wooldridge, 2020, p. 75; Cunningham, 2021, p. 65). Lovell (1963) later extended the result to multiple predictors, which is now commonly referred to as the Frisch–Waugh–Lovell (FWL) theorem or the regression anatomy formula. In econometrics, this result is typically presented as the FWL theorem (regression anatomy), whereas in statistics it is often framed as an orthogonal projection result (or orthogonal projection; Stapleton, 2009, p. 29; Seber & Lee, 2012, p. 475).

A standard way to understand the FWL theorem is through an equivalent two-step calculation based on observed variables. First, project (regress) the treatment assignment Ton the covariate $X$and retain the residual, $\tilde{T}$, which is the component of Torthogonal to X. Second, regress Yon this residual $\tilde{T}$. This two-step procedure yields exactly the same population-level ANCOVA

estimator for the coefficient on Tas the original ANCOVA regression that includes X directly. This can be written compactly as:

$$\gamma_{\text{ANCOVA}} = \frac{\text{Cov}\,(Y, \tilde{T})}{\text{Var}\,(\tilde{T})}, \qquad \tilde{T} = T - \frac{\text{Cov}\,(T, X)}{\text{Var}\,(X)} X.$$

The key point of using FWL theorem is that ANCOVA's population-level treatment coefficient is defined using $\tilde{T}$, so any remaining confounding association that survives after conditioning on X enters the ANCOVA coefficient through (1) the remaining association with Y in the residualized space and (2) the normalization by $\text{Var}\,(\tilde{T})$. This framing is important for linking the tracing-based graphical decomposition which identifies the remaining open backdoor confounding component after conditioning on X to the exact population OLS characterization of structural bias for ANCOVA, where that confounding component can be multiplied by residualization-based variance/covariance terms.

Although this use of the FWL theorem differs from its usual two-step presentation, the difference is algebraic rather than causal. In particular, some residualization steps may regress a causally upstream variable, such as the unobserved confounder A representing latent ability, on a causally downstream variable, such as the pretest X, in a direction opposite to the arrows in the causal graph. This does not create a contradiction, because the FWL theorem is a projection result: it decomposes a variable into the part explained by the conditioning variable and the residual part orthogonal to it, regardless of whether the variable being residualized is causally upstream or downstream. One potential source of confusion is that, in later derivations, $\tilde{T}$ may be algebraically re-expressed using projection identities implied by the assumed data-generating model, and it sometimes yielding expressions that involve residualized confounders (e.g., $\tilde{A}$) rather than only the observed covariate X. This can look "reversed" relative to the familiar two-step presentation,

but the logic is the same: $\tilde{T}$ is always defined as the part of T orthogonal to X, and any additional substitutions simply use model-implied equalities to rewrite that same residual in a form that is more interpretable for structural-bias decomposition. For example, when an unobserved confounder A is correlated with X, it is often useful to define the residualized confounder

$$\tilde{A} = A - \frac{\text{Cov}\,(A, X)}{\text{Var}\,(X)} X,$$

and then express the residualized treatment $\tilde{T}$ in terms of $\tilde{A}$ to make explicit how the confounding component remaining after conditioning on X is carried into the ANCOVA coefficient and how it is scaled by residual variation after partialling out X. This is precisely the role that the FWL theorem plays in deriving the structural bias results presented in the following sections.

3.2. Scenarios of DAGs for Two-wave Data

This section explores various scenarios of Directed Acyclic Graphs (DAGs) pertinent to two-wave data, utilizing the methods for identifying bias discussed earlier. By analyzing different relationships between pretest and posttest variables, along with the presence of confounders both observed and unobserved, I aim to illustrate how these factors contribute to bias in estimating the treatment effect. The forthcoming subchapters examine specific cases where the pretest does not affect the posttest, directly affects it, or affects it both directly and indirectly through a mediator. Additionally, I analyze scenarios involving single and multiple unobserved confounders, considering whether they are correlated or uncorrelated, and how their relationships impact the estimation of treatment effects. Utilizing ANCOVA and Change score analytic methods, I identify potential bias in each DAG scenario which means each data-generating model with different assumptions for structural causal relationships between the variables. This examination establishes a framework for addressing the complexities of potential bias in estimating the treatment effect

using two-wave data and emphasizes the necessity of appropriately controlling for confounding variables in research design.

Each hypothetical scenario comprises distinct cases that illustrate different dynamics between the pretest and the posttest scores, using "honors math program" as our running example. Based on recent national-level reports from U.S. principals in elementary and middle schools, assessment-based placement accounted for 7 percent at 5th-grade mathematics classes assignments, but it increased to almost 40 percent at 6th-grade, indicating increased placement-testing for prespecified path (e.g., honors math or career tracks) at middle school entry (Kaufman et al., 2024, p. 23). I assume that students who exceed a specified cutoff of the placement test (i.e., pretest) enter the honors math program (i.e., treatment group), while all others remain in the standard math curriculum (i.e., control group). I also assume the achievement test (i.e., posttest), administered one year later, mirrors the pretest in terms of format, enabling direct score comparisons and the computation of a change score (i.e., posttest minus pretest).

I analyze three scenarios distinguished by the number of unobserved confounders in a two-wave nonequivalent control-group design: Scenario 1 assumes no other unobserved confounder[8]; Scenario 2 introduces one other unobserved confounder A (e.g., student's math ability) that affects pretest X, treatment T (e.g., honors math program), and posttest Y; Scenario 3 adds a second other unobserved confounder B (e.g., parents' math ability), in addition to the unobserved confounder A.

In Scenario 1, no unobserved confounders are included. Case A is the ideal no-confounding baseline case; the pretest X is the only cause of treatment T and has no direct effect on the posttest

[8] I use the term "other" because the pretest X can itself confound the treatment–outcome relation (when $X \rightarrow T$ and $X \rightarrow Y$). Since X is observed and included in structural causal model, "other unobserved confounders" refers to additional, unmeasured causes of both T and Y beyond X (e.g., *A and B*).

Y. The DAG contains $X \rightarrow T$ and $T \rightarrow Y$ but no $X \rightarrow Y$, so placement on X does not confound the causal effect of T on Y. In Case B, the pretest score X is assumed as a stock variable representing math ability that persists from the pretest to the posttest. Therefore, the DAG includes a causal edge $X \rightarrow Y$ while retaining $X \rightarrow T$. Because X affects both the treatment T and the posttest Y, it acts as an observed confounder of the treatment effect $T \rightarrow Y$.

Scenario 2 introduces a single unobserved confounder A, a math ability. In Case A, I continue to assume the pretest X as a stock variable representing math ability. Accordingly, once the unobserved math ability A is included in the DAG, X serves only as proxy for A. The direct edge $X \rightarrow Y$ is therefore eliminated, since the X–Y association is attributed to their common cause A in this case. To allow X to affect Y beyond what A explains, Case B adopts the pretest as a mixture perspective. In this case, X still serves as a possibly imperfect proxy for A via the path $A \rightarrow X$, but it may also reflect other stock-like components (e.g., test taking proficiency) as well as flow-like components, that is, time specific states or processes that can change between waves and plausibly carry forward (e.g., motivational responses when pretest information becomes salient to students). The DAG therefore retains a direct edge $X \rightarrow Y$ to represent residual carryover from the pretest that is not fully captured by stable ability A. Building on the pretest as a mixture perspective, Case B leaves the specific mechanism behind the residual $X \rightarrow Y$ carryover unspecified, whereas Case C makes one plausible between wave mechanism explicit by introducing a mediator M. For example, once students learn their pretest score, that information can shape self-efficacy, which can then influence subsequent effort, engagement, and learning, ultimately affecting the posttest outcome. In this way, Case C extends Case B by adding the path $X \rightarrow M \rightarrow Y$ while retaining the direct effect $X \rightarrow Y$, thereby formalizing a motivational response mechanism that Kim and Steiner (2021) discuss informally but do not model with an explicit

mediator. I kept the direct edge alongside the mediation path because complete mediation is uncommon in practice: a single mediator such as self-efficacy is likely to explain only part of the total pretest to posttest influence, and eliminating the direct effect would typically require specifying additional mediating processes.

Scenario 3 introduces an additional unobserved confounder, B, representing parents' math ability, in addition to the student-level ability confounder A. Substantively, B can influence students' baseline achievement X (for example, the quality of math help they can provide at home, or the selection of more mathematically enriching activities) and can also affect treatment assignment T (for example, evaluation of program options, communicate effectively with teachers about math placement, or advocate for advanced coursework). The three cases in Scenario 3 follow the same conceptual progression as under Scenario 2, differing only in how the pretest is interpreted. In Case A, X is treated as a pure stock proxy for A. Once A is included in the DAG, any persistence from pretest to posttest is attributed to the stable trait A, so the direct edge $X \rightarrow Y$ is omitted. In Case B, X still functions as a proxy for A, but under the pretest as a mixture perspective it can also contain additional carryover that is not fully captured by A, such as test taking proficiency or motivational responses triggered by knowing the pretest score; therefore, the direct edge $X \rightarrow Y$is retained. In Case C, the between wave mechanism is made more explicit by adding an unobserved mediator M (for example, self-efficacy), yielding the path $X \rightarrow M \rightarrow Y$ while keeping $X \rightarrow Y$ to allow for partial mediation. Because M is unobserved and is assumed to be driven primarily by prior achievement X and treatment T between waves, any influence of parental math ability Bon M can be represented through the indirect paths $B \rightarrow X \rightarrow M$ and $B \rightarrow T \rightarrow M$. I therefore omitted a direct $B \rightarrow M$ edge, since with only two waves and an unobserved mediator it would be difficult to distinguish a direct effect of B on M from B's overall unobserved

influence on the posttest outcome Y. Alongside this conceptual distinction, Scenario 3 treats A and B as substantively distinct sources of unobserved confounding rather than as interchangeable latent variables. This asymmetry clarifies how student level stable ability versus family level influences enter the open backdoor structure, and it makes the ANCOVA versus change score bias comparison more informative.

In structural causal models across all scenarios and cases, I assume that causal effects are linear and additive, that interaction effects are absent, and that error terms are mutually independent random variables. I also assume that all variables except Y are measured without error, and standardized (i.e., mean-centered with a unit variance). The random errors (i. e., $\varepsilon_X, \varepsilon_T, \varepsilon_M,$ and ε_Y) are not explicitly represented in the DAGs, which is usually the convention for DAGs. In line with Lüdtke and Robitzsch (2025, p. 365), Y is not standardized, and its variance can differ from one, which accommodates individual differences in growth over time (i.e., the "fan spread effect"; Campbell & Kenny, 1999; Kenny, 1975). In DAGs underlying all scenarios and cases, filled nodes denote observed variables, unfilled nodes denote unobserved variables, solid directed edges indicate relations among observed variables, and dashed directed edges indicate relations involving unobserved variables.

3.2.1. Scenario 1: No Unobserved Confounders

In this hypothetical scenario, I assume that no confounding variables cause changes in pretest and posttest scores or the treatment assignment, ensuring that unobserved confounders do not bias any relationships among these variables. Each DAG under three distinct cases represents the graphical model of structural causal model (i.e., data-generating model) used in ANCOVA or change score analytic method having observed variables pretest scores X, treatment T, and posttest scores Y.

3.2.1.1. Case A: When the pretest does not affect the posttest

In this idealized baseline Scenario1: Case A, I assume that pretest scores do not directly cause any changes in posttest scores for individual participants. In other words, the pretest score functions solely as a sorting variable for program assignment, and any subsequent change in posttest performance is attributed entirely to the treatment or to random variation. Specifically, at the end of fifth grade students complete a mathematics achievement test (i.e., pretest): those above the cutoff enter the honors math program (i.e., treatment group), while those below remain in the standard curriculum (i.e., control group), and at the end of sixth grade they undertake a mathematics achievement test again (i.e., posttest) based on which honors students below the retention cutoff assign to the standard curriculum and those above it remain in the honors program. Under case A, simply observing one's pretest result exerts no causal influence on the posttest score. Although earlier scores often provide some indication of later performance in real classrooms, adopting this "no direct effect" baseline offers a clear reference case before I move on to more plausible cases in which including the pretest to posttest edge alters the identification of structural bias in treatment effect.

Suppose the variable X represents a pretest scores of math (i.e., baseline scores), T denotes the binary treatment variable indicating placement in the Honors Math Program versus the standard program, and Y is the posttest math scores (i.e., the outcome measured after the placement). The data-generating model, or the structural causal model under Scenario 1: Case A can be specified as follows:

$$T = \delta_T X + \varepsilon_T,$$
$$Y = \gamma T + \varepsilon_Y.$$

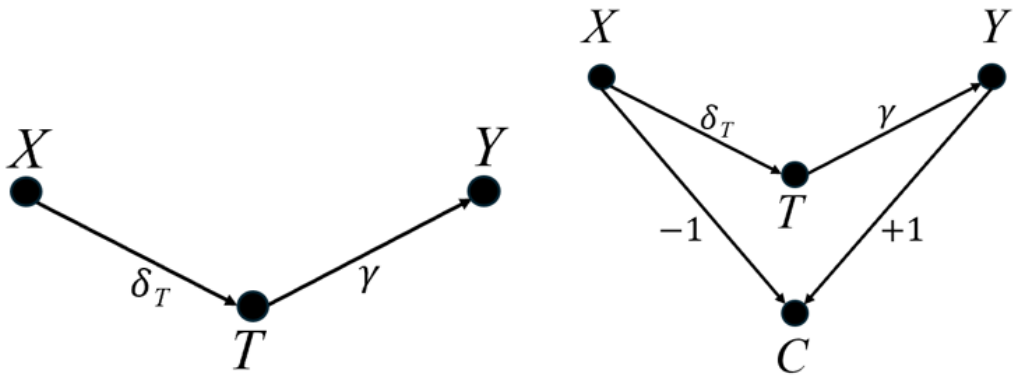

Figure 3. Scenario 1, Case A: no unobserved confounders and no causal effect of the pretest on the posttest.

The path coefficient γ shows the true treatment effect on the posttest, and the non-zero path coefficient δ_T captures the effect of the pretest on treatment assignment assuming the non-equivalent comparison groups without random assignment. Figure 3 presents two DAGs based on the same underlying data-generating model for Scenario 1, Case A: the left DAG corresponds to ANCOVA, and the right DAG corresponds to change score analysis. Thus, the difference between the two DAGs does not reflect different causal structures, but different analytic representations of the outcome. In the ANCOVA DAG, the outcome is the posttest Y itself. In the change score analysis DAG, the outcome is the computed variable $C = Y - X$ rather than Y. Accordingly, the change score analysis DAG includes C and represents its construction from Y and X using fixed coefficients of +1 and -1 (Pearl, 2016; Kim & Steiner, 2021). These arrows do not indicate causal effects of Y or X on C; rather, they indicate that C is deterministically calculated from the observed scores Y and X. Representing C in this way is useful because it makes the change-score outcome explicit within the DAG and facilitates the application of the backdoor criterion to trace confounding paths in the change score analysis model. In particular, including C clarifies that the analysis does not condition on X or Y separately, but instead uses their linear combination as the outcome. The same logic applies to the corresponding DAGs in the subsequent scenarios as well.

3.2.1.2. Case B: When the pretest affects the posttest

This case assumes that pretest score *X* directly cause changes in posttest score *Y*, unlike the previous case, which did not make this causal path. In other words, students who participated in either the honors math program or the regular math curriculum, the pretest math achievement score is assumed to directly cause changes in posttest math scores, not in the way of "predictive" reason. If I view the pretest score *X* as a stock variable representing math ability that inherently persists over time, it follows that this ability carries over from the pretest to the posttest period. Therefore, I can assume a causal effect of the pretest score on the posttest score, as the initial level of math ability directly influences subsequent math ability. This perspective supports the notion of a direct causal relationship between pretest and posttest scores in this context.

The DAGs for Case B included the directed edge from *X* to *Y* with the path coefficient *τ,* to the DAGs of Case A (see Figure 4). Accordingly, the data-generating model under Scenario 1: Case B can be specified as follows:

$$T = \delta_T X + \varepsilon_T,$$
$$Y = \tau X + \gamma T + \varepsilon_Y.$$

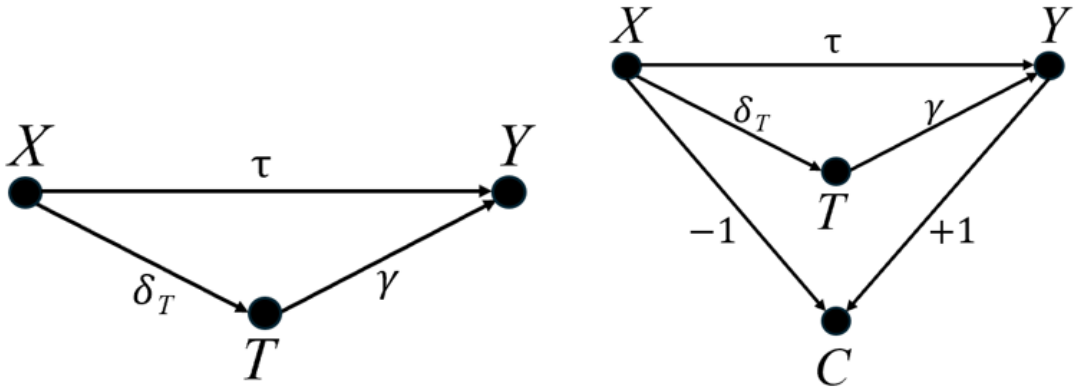

Figure 4. Scenario 1, Case B: no unobserved confounders and a causal effect of the pretest on the posttest.

3.2.2. Scenario 2: Single Unobserved Confounder

Under Scenario 2, I extend Scenario 1 by introducing an unobserved confounder A, each student's math ability, which causally influences the pretest score X, the posttest score Y, and placement in the honors math program T. Scenario 2 comprises three distinct cases that illustrate different dynamics between pretest and posttest measures when participants are assigned to treatment or control based on their observed pretest scores and unobserved math ability. After identifying structural bias in both the ANCOVA and Change score analytic methods in the presence of an unobserved confounder A, I compare how those biases differ between the two analytic methods.

Each DAG under three distinct cases represents the graphical model of structural causal model used in ANCOVA or change score analytic method having observed variables pretest scores X, treatment T, and posttest scores Y, as well as an unobserved confounder such as a student's math ability A.

3.2.2.1. Case A: When the pretest does not affect the posttest

This case assumes the pretest scores X do not directly cause changes in posttest scores Y. Under Scenario 1: Case B, I treated pretest X as a stock variable that carries over the math ability to justify the causal effect of X on Y when I did not specify the latent variable of student's math ability. Now under Scenario 2, I posit a consistent, time-invariant student's math ability A that removes the causal effect of X on Y as a stock variable. Since the math ability A consistently influences X and Y, any direct causal effect from the pretest score X to the posttest score Y is completely explained by their shared cause A. The test scores are simply "indicator" or "index" of latent variable, unobserved math ability A, which is treated here as a stock variable. In other words, X acts only as a possibly imperfect "proxy" for A in econometrics term (Kim & Steiner, 2021).

Once A is included, X no longer carries an independent causal effect on Y as it did when A was omitted in Scenario 1, Case B. Therefore, the omission of the directed edge $X \rightarrow Y$ is justified, leaving only the dashed directed edges $A \rightarrow X$ and $A \rightarrow Y$. In Scenario 2, I also assume that meeting the pretest cutoff is necessary but not sufficient for honors program placement because final decisions also consider broader criteria (e.g., cumulative grades, teacher recommendations). Accordingly, I include a direct directed edge $A \rightarrow T$ to reflect that a student's math ability can influence honors program placement.

Based on the DAGs of Scenario 1: Case A, the DAGs of Scenario 2: Case A included an unobserved confounder A (e.g., student's math ability) having directed edges to pretest scores X, posttest scores Y, and treatment assignment T with the path coefficients $\alpha_1, \alpha_2, and\ \alpha_T$ respectively (see Figure 5). The data-generating model under Scenario 2: Case A can be specified as follows:

$$X = \alpha_1 A + \varepsilon_X,$$
$$T = \alpha_T A + \delta_T X + \varepsilon_T,$$
$$Y = \alpha_2 A + \gamma T + \varepsilon_Y.$$

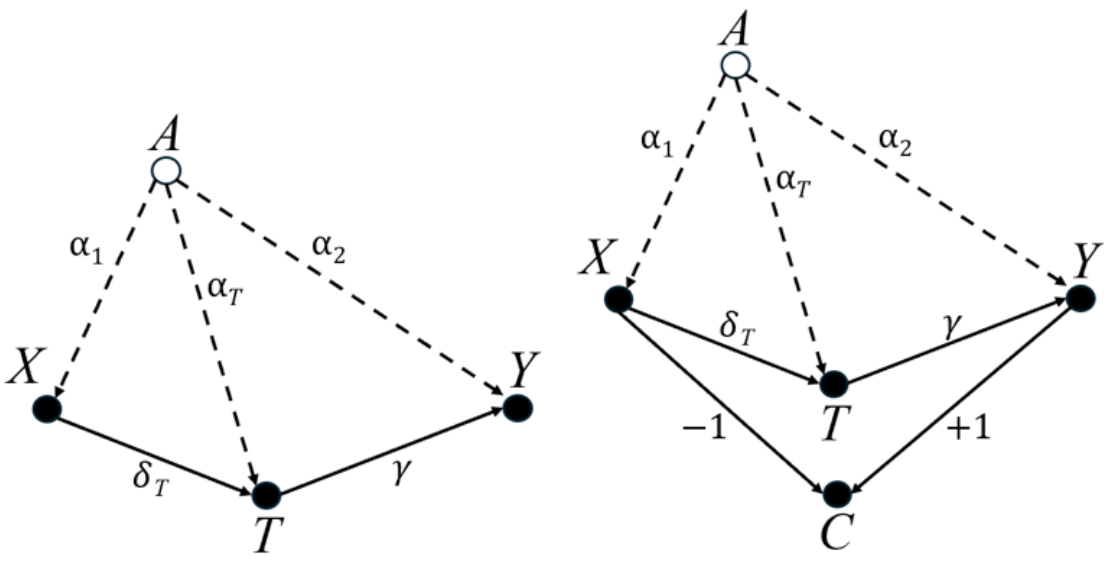

Figure 5. Scenario 2, Case A: one unobserved confounder and no causal effect of the pretest on the posttest.

3.2.2.2. Case B: When the pretest affects the posttest

In this case I allow Grade 5 pretest score X to directly affect the Grade 6 posttest score Y even after I account for math ability A. Because real-world test scores, whether cognitive or attitudinal, rarely represents a single "pure" indicator or index of a single latent variable, it is reasonable to treat pretest X as a "mixture" of stock and flow components in proportions that cannot be observed. In this sense, a Grade 5 math pretest is therefore more than a pure indicator or proxy for the single stock variable of math ability A; it may capture flow-like component such as current interest or motivation in mathematics that can be reinforced to affect later outcome Y.

Since the exact stocks-versus-flows ratio in X is unknown, it is reasonable to treat it as a mixture of both. This logic justifies retaining the direct arrow $X \rightarrow Y$ after controlling for the math ability A in the DAG; Part of X is determined by the math ability A, yet the remainders still affect changes in the posttest Y. The DAGs included a directed edge from pretest X to posttest Y with the path coefficient τ, to the DAGs of Case A (see Figure 6). The data-generating model under Scenario 2: Case B can be specified as follows:

$$X = \alpha_1 A + \varepsilon_X,$$
$$T = \alpha_T A + \delta_T X + \varepsilon_T,$$
$$Y = \alpha_2 A + \tau X + \gamma T + \varepsilon_Y.$$

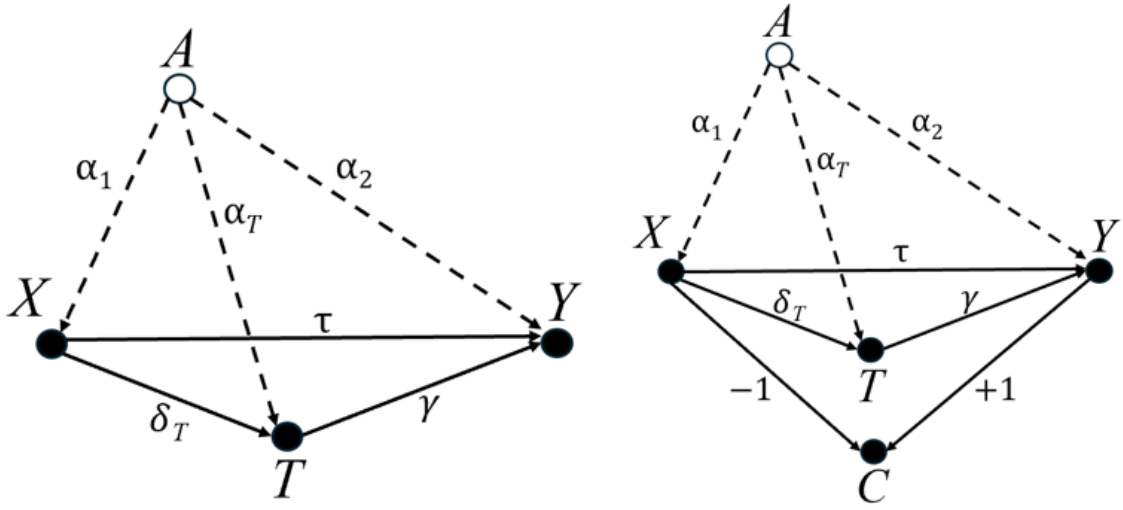

Figure 6. Scenario 2, Case B: one unobserved confounder and a causal effect of the pretest on the posttest.

3.2.2.3. Case C: When the pretest affects the posttest through a mediator

Case C extends Case B by retaining the direct effect $X \rightarrow Y$ while also introducing an explicit mediating path $X \rightarrow M \rightarrow Y$, where M represents a between-wave psychological process such as math self-efficacy. Consistent with the pretest as a mixture perspective, X reflects not only stable, stock-like components of achievement but can also trigger newly generated, flow-like responses: for example, learning one's pretest result may update self-efficacy, which then shapes subsequent effort, engagement, and learning and thereby influences the posttest. In this way, Case C makes the reinforcing carryover mechanism between waves explicit while allowing for partial mediation by keeping $X \rightarrow Y$ since complete mediation by a single process is unlikely in practice. In this specification, X now affects Y both directly (path coefficient τ) and indirectly through a student's unobserved math self-efficacy M. Case C retains the unobserved mediator M; it receives

arrows from X (δ_M) and from T (ζ) and carries its effect to Y (η). The unobserved confounder A (i.e., math ability) continues to affect X (α_1), T (α_T), and Y (α_2), and it also affects M (α_M) . The data-generating model for Scenario 2: Case C is therefore:

$$X = \alpha_1 A + \varepsilon_X,$$
$$T = \alpha_T A + \delta_T X + \varepsilon_T,$$
$$M = \alpha_M A + \delta_M X + \zeta T + +\varepsilon_M,$$
$$Y = \alpha_2 A + \tau X + \gamma T + \eta M + \varepsilon_Y.$$

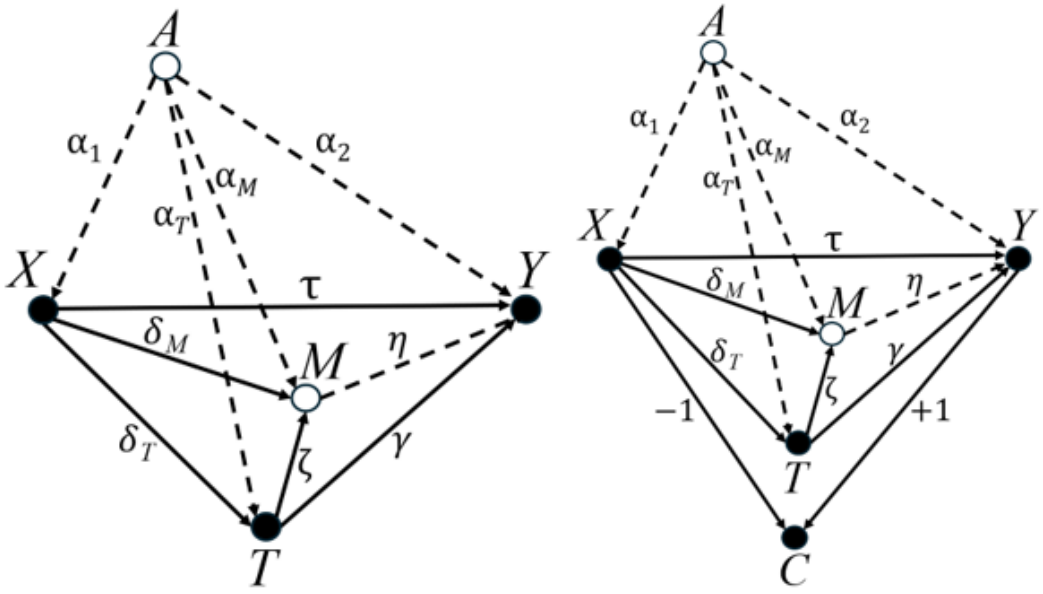

Figure 7. Scenario 2, Case C: one unobserved confounder and causal effects of the pretest on the posttest through both direct and mediated paths.

3.2.3. Scenario 3: Two Unobserved Confounders

In this scenario, I added one more confounding variable *B* to scenario 2. For example, another confounder *B* is parents' math ability, which causes changes in pretest scores, posttest scores, and treatment assignment in the context of honors math program example. I assume an unobserved common cause *U* representing the family's genetic endowment, which causally affects both the parent's math ability *B* and the student's math ability *A*, thereby inducing a correlation

between A and B.[9] Therefore, the two confounders A and B are correlated by receiving the directed edges from the unobserved common cause U with the coefficients $\rho_A \; and \; \rho_B$ respectively: $A = \rho_A U + \varepsilon_A, B = \rho_B U + \varepsilon_B$. Both confounders have directed edges to pretest scores, posttest scores, and treatment assignment. This correlation between two unobserved confounders implies additional backdoor paths to add bias for the treatment effect.

Applying the same rationales for each causal path as in Scenario 2, Scenario 3 is likewise divided into three cases: A, B, and C. Under the DAGs of each case, I conduct ANCOVA and Change score analyses to identify the bias for the treatment effect. The differences in identified biases between the analytic models will be compared.

3.2.3.1. Case A: When the pretest does not affect the posttest

This case assumes the pretest scores X do not directly cause changes in posttest scores Y. Based on the previous DAGs under Scenario 2, The DAGs in Scenario 3 added the unobserved confounder B having directed edges to pretest scores, posttest scores, and treatment assignment with the coefficients $\beta_1, \beta_2, and \; \beta_T$ respectively (see Figure 8). The data-generating model under Case A is specified as follows:

$$X = \alpha_1 A + \beta_1 B + \varepsilon_X,$$
$$T = \alpha_T A + \beta_T B + \delta_T X + \varepsilon_T,$$
$$Y = \alpha_2 A + \beta_2 B + \gamma T + \varepsilon_Y.$$

[9] According to Morgan & Winship (2014, p. 80-81), a correlation between two variables in a DAG indicates their shared dependence on an unobserved common cause. In a DAG, which adheres to the assumption of acyclicity, relationships should be represented using directed edges rather than bidirectional arrows. Therefore, instead of using a bidirectional arrow between two variables, it is more appropriate to represent their relationship by drawing a common unobserved variable with directed arrows pointing to both variables. This representation makes it clear that the relationship between the two variables arises from a shared unobserved factor and ensures that the diagram remains acyclic.

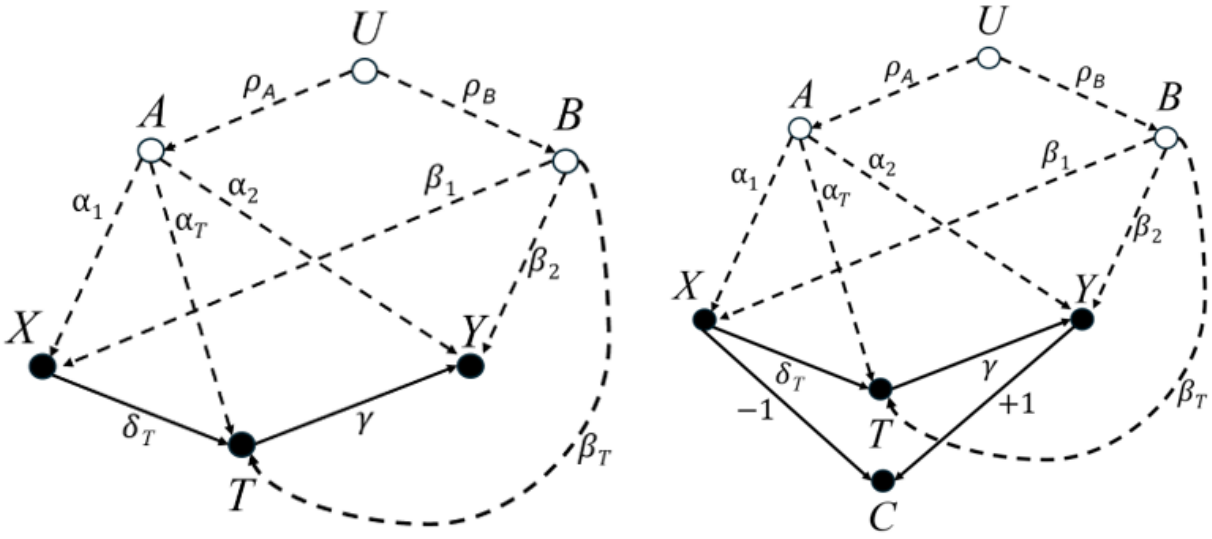

Figure 8. Scenario 3, Case A: two unobserved confounders and no causal effect of the pretest on the posttest

3.2.3.2. Case B: When the pretest affects the posttest

This case assumes that the pretest score *X* directly affects the posttest score *Y*. The DAGs added a directed edge from the pretest to the posttest with the coefficient τ (see Figure 9). The data-generating model under Case B is specified as follows:

$$X = \alpha_1 A + \beta_1 B + \varepsilon_X,$$
$$T = \alpha_T A + \beta_T B + \delta_T X + \varepsilon_T,$$
$$Y = \alpha_2 A + \beta_2 B + \tau X + \gamma T + \varepsilon_Y.$$

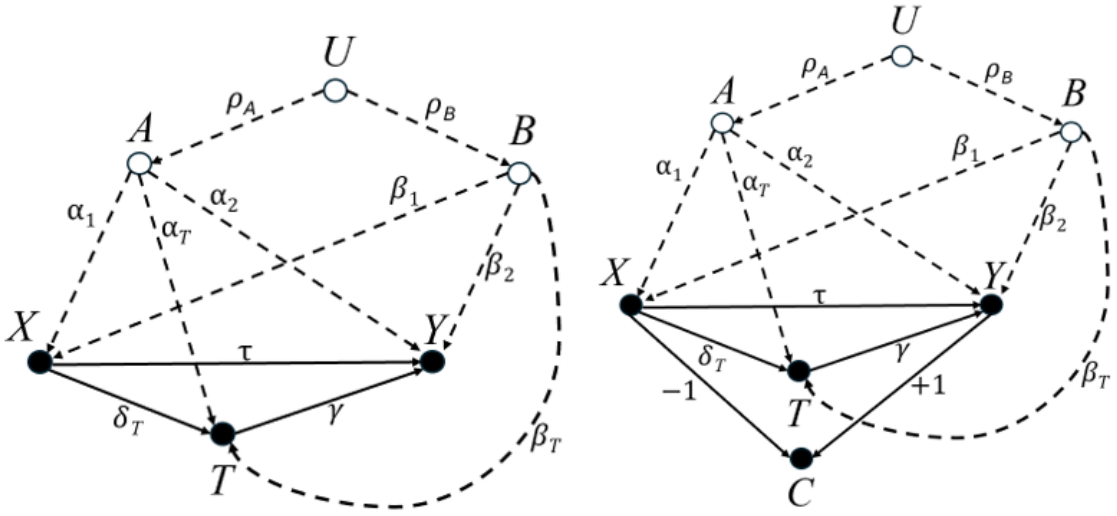

Figure 9. Scenario 3, Case B: two unobserved confounders and a causal effect of the pretest on the posttest

3.2.3.3. Case C: When the pretest affects the posttest through a mediator

Following the same logic of Scenario 2, Case C, this case added the indirect path where the pretest score X indirectly affects the posttest score Y through an unobserved mediating variable M (i.e., a student's math self-efficacy). The DAGs for Case C introduced the mediating variable M, which received a directed edge from X with the coefficient δ_M, and sent a directed edge to Y with the coefficients η (see Figure 10). I conceptualize M(e.g., self-efficacy) as a between-wave, within-student mechanism that is caused by performance and program exposure. Accordingly, I include $X \rightarrow M$ and $T \rightarrow M$, and allow $A \rightarrow M$ to capture this mechanism in such states. Parents' math ability B is modeled as operating on M chiefly through its earlier effects on $X$and access to T; with only two waves, any residual direct $B \rightarrow M$ is not separately identifiable and is absorbed into B's total effect on Y. I therefore omit $B \rightarrow M$ without loss of generality. The data-generating model in scenario 3.1: Case C can be specified as follows:

$$X = \alpha_1 A + \beta_1 B + \varepsilon_X,$$
$$T = \alpha_T A + \beta_T B + \delta_T X + \varepsilon_T,$$
$$M = \alpha_M A + \delta_M X + \zeta\, T + \varepsilon_M,$$
$$Y = \alpha_2 A + \beta_2 B + \tau\, X + \gamma\, T + \eta\, M + \varepsilon_Y.$$

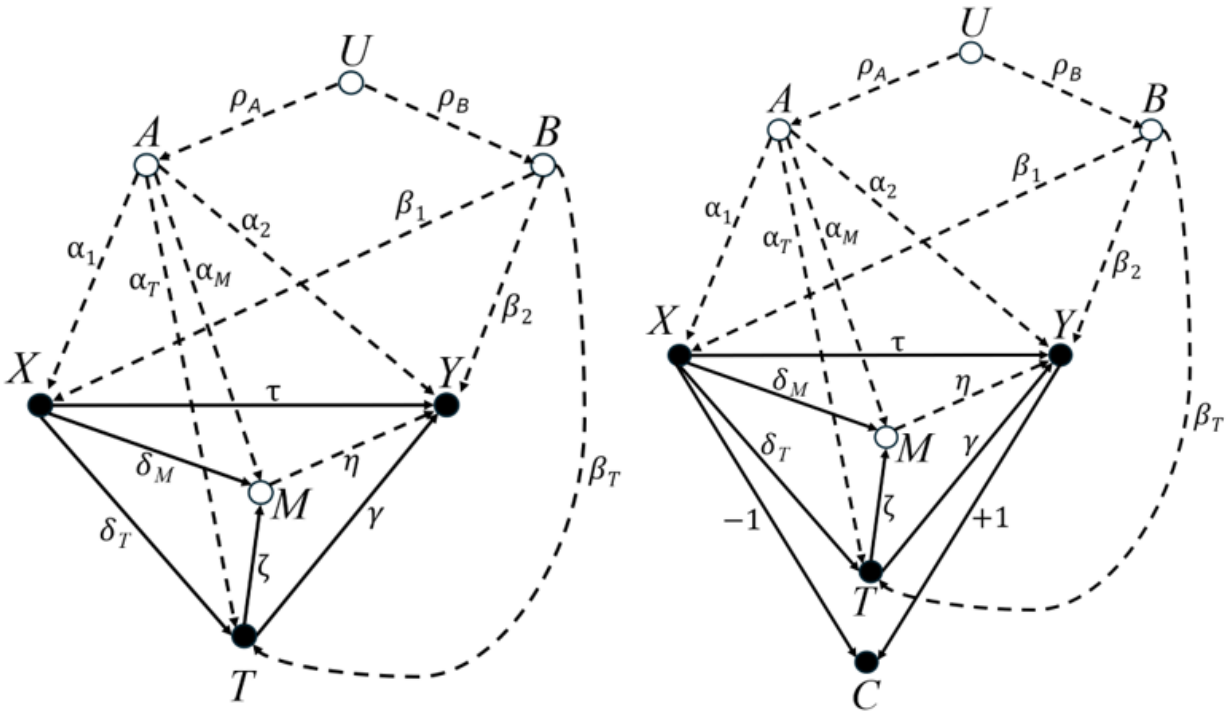

Figure 10. Scenario 3, Case C: two unobserved confounders and causal effects of the pretest on the posttest through both direct and mediated paths

4. Results

This chapter presents the results of the study. The study aimed to compare the structural bias and sufficient unbiasedness conditions of ANCOVA and change-score analysis in two-wave nonequivalent control-group designs and to examine how the graphical results correspond to the population-level OLS results. The chapter is organized by scenario. Section 4.1 presents results for Scenario 1, which assumes no unobserved confounders; Section 4.2 presents results for Scenario 2, which includes a single unobserved confounder; and Section 4.3 presents results for Scenario 3, which includes two unobserved confounders. Within each scenario, results are reported by case, and each case presents four sets of results: ANCOVA using the graphical approach, ANCOVA using the Frisch–Waugh–Lovell theorem, change-score analysis using the graphical approach, and change-score analysis using population-level OLS.

4.1. Scenario 1: No Unobserved Confounders

4.1.1. Case A: When the pretest does not affect the posttest

4.1.1.1. ANCOVA: Graphical Approach

Under Scenario 1, Case A, there are no open backdoor paths from T to Y for the ANCOVA estimator in Equation (1) which is the population partial regression coefficient of Y on T controlling for X, denoted γ_{ANCOVA}. Since the DAG contains only the directed causal chain $X \rightarrow T \rightarrow Y$, there are no spurious associations through which *T* might cause *Y*, implying that the causal effect of *T* on *Y*, the treatment effect, is not biased for ANCOVA estimator. Let $Bias(\gamma_{ANCOVA,1A})_G$ denotes the structural bias identified by the graphical approach for the ANCOVA estimator under Scenario 1, Case A. For consistency, bias expressions obtained from the graphical approach are marked with the subscript G throughout the dissertation. Then,

$$\boldsymbol{Bias}(\boldsymbol{\gamma_{ANCOVA,1A}})_{\boldsymbol{G}} = \mathbf{0}.$$

4.1.1.2. ANCOVA: OLS using the FWL theorem

By the zero conditional mean assumption,

$$\mathrm{Cov}(\varepsilon_T, X) = 0, \mathrm{Cov}(\varepsilon_T, \varepsilon_Y) = 0.$$

1) Residualization

Let $\gamma_{\mathrm{ANCOVA,1A}}$ denote the ANCOVA estimator under Scenario 1, Case A. By the FWL theorem or regression anatomy, the ANCOVA estimator is defined as

$$\gamma_{ANCOVA,1A} = \frac{\mathrm{Cov}(Y, \tilde{T})}{\mathrm{Var}(\tilde{T})},$$

where $\tilde{T}$ denotes the residual of T after partialling out X:

$$\tilde{T} = T - \frac{\mathrm{Cov}(T, X)}{\mathrm{Var}(X)} X.$$

I first define the fitted value

$$\hat{T} = bX$$

from projecting T on X, where b is the population OLS slope, then

$$\tilde{T} = T - \hat{T}.$$

Here, I work with mean centered variables,

$$\mathbb{E}(T) = \mathbb{E}(X) = 0,$$

so that $\hat{T} = bX$ without an intercept.

Then I can define

$$\tilde{T} = T - bX$$

with b chosen so that

$$\mathrm{Cov}(\tilde{T}, X) = 0$$

by the OLS orthogonality condition.

Hence,

$$\mathrm{Cov}(T - bX, X) = \mathrm{Cov}(T, X) - b\mathrm{Var}(X),$$

then

$$b = \frac{\mathrm{Cov}(T, X)}{\mathrm{Var}(X)}.$$

2) Compute $\tilde{T}$

From the $T = \delta_T X + \varepsilon_T$,

$$\mathrm{Cov}(T, X) = \delta_T \mathrm{Var}(X) + \mathrm{Cov}(\varepsilon_T, X) = \delta_T \mathrm{Var}(X).$$

So

$$\tilde{T} = T - \delta_T X = \varepsilon_T.$$

Thus, the residual $\tilde{T}$ is equal to ε_T under Scenario 1: Case A.

Hence, by the data-generating model or structural causal model (SCM),

$$\mathrm{Cov}\left(X, \tilde{T}\right) = \mathrm{Cov}(X, \varepsilon_T) = 0,$$

$$\mathrm{Var}\left(\tilde{T}\right) = \mathrm{Var}(\varepsilon_T).$$

3) Compute numerator: $\mathrm{Cov}\left(Y, \tilde{T}\right)$

$$\begin{aligned} \mathrm{Cov}\left(Y, \tilde{T}\right) &= \mathrm{Cov}\left(\gamma T + \varepsilon_Y, \tilde{T}\right) \\ &= \gamma \mathrm{Cov}\left(T, \tilde{T}\right) + \mathrm{Cov}\left(\varepsilon_Y, \tilde{T}\right) \\ &= \gamma \mathrm{Cov}\left(T, \tilde{T}\right). \end{aligned}$$

$$\begin{aligned} \mathrm{Cov}\left(T, \tilde{T}\right) &= \mathrm{Cov}(\delta_T X + \varepsilon_T, \varepsilon_T) \\ &= \delta_T \mathrm{Cov}(X, \varepsilon_T) + \mathrm{Var}(\varepsilon_T). \end{aligned}$$

Thus,

$$\mathrm{Cov}\left(Y, \tilde{T}\right) = \gamma \mathrm{Var}(\varepsilon_T).$$

4) Compute denominator: $\mathrm{Var}\left(\tilde{T}\right)$

$$\mathrm{Var}\left(\tilde{T}\right) = \mathrm{Var}(\varepsilon_T).$$

5) Accordingly,

$$\boldsymbol{\gamma_{ANCOVA,1A}} = \frac{\mathrm{Cov}(Y, \tilde{T})}{\mathrm{Var}(\tilde{T})} = \gamma \cdot \frac{\mathrm{Var}(\varepsilon_T)}{\mathrm{Var}(\varepsilon_T)} = \gamma,$$

which means no structural bias for ANCOVA estimator under Scenario 1, Case A. Let $Bias(\gamma_{ANCOVA,1A})_{OLS}$ denotes the structural bias identified by the OLS for the ANCOVA estimator under Scenario 1, Case A. For consistency, bias expressions obtained from the OLS are marked with the subscript OLS throughout the dissertation. Then, $\boldsymbol{Bias(\gamma_{ANCOVA,1A})_{OLS} = 0}$.

Therefore, $Bias(\gamma_{ANCOVA,1A})_G$ is identical to $Bias(\gamma_{ANCOVA,1A})_{OLS}$ which is zero, when the data-generating model contains no unobserved confounders and no direct pretest to posttest effect. The only causal path from treatment to outcome is $T \rightarrow Y$, so there are no open backdoor paths between $T$and Y. Consequently, conditioning on X in ANCOVA is not required for treatment effect identification, but it is also harmless: the ANCOVA estimator equals the causal treatment effect, so the structural bias is zero. The graphical approach reaches this conclusion immediately because the DAG contains no backdoor path to block. The OLS result via the FWL theorem is consistent with the same logic: residualizing Ton X yields $\tilde{T} = \varepsilon_T$, and because $Y = \gamma T + \varepsilon_Y$, $\mathrm{Cov}\,(Y, \tilde{T}) = \gamma\,\mathrm{Var}\,(\tilde{T})$ under the usual orthogonality assumptions, implying $\gamma_{ANCOVA,1A} = \gamma$.

4.1.1.3. Change Score Analysis: Graphical Approach

Under Scenario 1, Case A, there is only one 'open' backdoor path from *T* to *C* for the change score estimator in Equation (2) which is the population regression coefficient of C on T denoted γ_{Change}:

$$\boldsymbol{T \leftarrow X \rightarrow C}.$$

This backdoor path is open because *X* is not conditioned upon in the change score analytic model. Unlike in the ANCOVA, the treatment effect obtained from the change score analysis

would be biased due to this unblocked backdoor path. As a result, the noncausal association via the open backdoor path is given by the product of the corresponding structural path coefficients on the path as a structural bias. Let $Bias(\gamma_{Change,1A})_G$ denotes the structural bias identified by graphical approach in the change score analysis under Scenario 1, Case A. Then, $\boldsymbol{Bias(\gamma_{Change,1A})_G} = \delta_T \times (-1) = \boldsymbol{-\delta_T}$.

4.1.1.4. Change Score Analysis: OLS

The change score analytic model is a simple regression model, and the change score estimator, γ_{Change}, is the population regression coefficient of C on T, defined as follows:

$$\gamma_{\text{Change}} = \frac{Cov(C,T)}{Var(T)}.$$

Substituting $C = Y - X$ into this expression:

$$\gamma_{\text{Change}} = \frac{Cov(Y-X,T)}{Var(T)}.$$

Now using the linearity of covariance for the numerator, and since T is standardized, $Var(T) = 1$. The final simplified expression for γ_{Change} is:

$$\gamma_{\text{Change}} = Cov(Y,T) - Cov(X,T).$$

Under Scenario 1, Case A, the covariances are given by:

$Cov(X,T)$[10] $= Cov(X, \delta_T X + \varepsilon_T) = \delta_T\, Var(X) + Cov(X, \varepsilon_T) = \delta_T \cdot 1 + 0 = \delta_T,$

$Cov(Y,T) = Cov(\gamma T + \varepsilon_Y, T) = \gamma\, Var(T) + Cov(\varepsilon_Y, T) = \gamma \cdot 1 + 0 = \gamma,$

[10] Since X and T are assumed to be standardized, I can express it differently such as, $\text{Cov}(X,T) = E[XT] = E[X(\delta_T X + \epsilon_T)] = \delta_T E[X^2] + E[X\epsilon_T]$. Since X is standardized, $E[X^2] = 1$, and because ϵ_T is independent of X, $E[X\epsilon_T] = 0$. By definition, $\text{Var}(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2$. Substituting $E[X] = 0$ and $\text{Var}(X) = 1$ into this definition gives $E[X^2] = 1$, which holds for any standardized random variable. Under the zero conditional mean assumption, I have $E[\,\epsilon_T \mid X\,] = 0$. Applying the law of iterated expectation, $E[X\varepsilon_T] = E\big[E[\,X\varepsilon_T \mid X\,]\big]$. I treat X as if it is a known constant within the conditional expectation because I am conditioning on the value that X has taken, thus, $E\big[E[\,X\varepsilon_T \mid X\,]\big] = E\big[X\,E[\,\epsilon_T \mid X\,]\big] = E[X \cdot 0] = 0$. Consequently, $E[X\varepsilon_T] = 0$. The same reasoning applies to other standardized random variables.

Note that $Cov(X, \varepsilon_T)$,[11] and $Cov(\varepsilon_Y, T)$[12] are equal to zero due to the zero conditional mean assumption (i.e., each error terms are uncorrelated to the independent variables and distributed with zero means), and the independence of the errors assumption (i.e., error terms are uncorrelated random variables). The same assumptions apply to all covariances between error terms and predictors in the data-generating models across scenarios. Let $\gamma_{Change,\ 1A}$ denotes the change score estimator under Scenario 1, Case A. Then, $\boldsymbol{\gamma_{Change,\ 1A}} = \gamma - \boldsymbol{\delta_T}$ where the structural bias under Scenario 1, Case A is: $-\delta_T$. Let $Bias(\gamma_{Change,1A})_{OLS}$ denotes the structural bias identified by the OLS for the change score estimator under Scenario 1, Case A. For consistency, bias expressions obtained from the OLS are marked with the subscript OLS throughout the dissertation. Then, $\boldsymbol{Bias(\gamma_{Change,1A})_{OLS}} = \boldsymbol{-\delta_T}$.

Therefore, $Bias(\gamma_{Change,1A})_G$ is identical to $Bias(\gamma_{Change,1A})_{OLS}$. The change score estimator is unbiased for the treatment effect under Scenario 1, Case A if and only if $\delta_T = 0$, meaning the pretest has no causal influence on treatment assignment in the data generating process. This unbiasedness condition is consistent with the requirement of "no dynamic causal relationships between the outcome and treatment variables" (Imai and Kim, 2019, p. 477; Lüdtke and Robitzsch, 2025) and is often difficult to satisfy in nonequivalent control group designs where treatment participation is frequently determined, at least in part, by baseline performance.

[11] Given the zero conditional mean assumption, $E[\,\varepsilon_T \mid X\,] = 0$. By the definition of covariance and applying the law of iterated expectation, $\text{Cov}(X, \varepsilon_T) = E[X\,\varepsilon_T] = E(E[\,X\,\varepsilon_T \mid X\,]) = E(X\,E[\,\varepsilon_T \mid X\,]) = 0$.

[12] Given the zero conditional mean assumption, $E[\,\varepsilon_Y \mid X, T\,] = 0$. By the definition of covariance and applying the law of iterated expectation, $\text{Cov}(X, \varepsilon_Y) = E[X\,\varepsilon_Y] = E(E[\,X\,\varepsilon_Y \mid X, T\,]) = E(X\,E[\,\varepsilon_Y \mid X, T\,]) = 0$.

4.1.2. Case B: When the pretest affects the posttest

4.1.2.1. ANCOVA: Graphical Approach

Under Scenario 1, Case B, the DAG implies a single potential backdoor path from T to Y induced by the pretest X:

$$T \leftarrow X \rightarrow Y.$$

However, ANCOVA conditions on X so that this path is blocked. Since there are no additional confounders, there are no other backdoor paths from T to Y that remain open after conditioning on X. Therefore, conditioning on X satisfies the backdoor criterion for identifying the causal effect of T on Y in this case, and the ANCOVA estimator is unbiased:

$$\boldsymbol{Bias}(\boldsymbol{\gamma_{ANCOVA,1B}})_{\boldsymbol{G}} = \mathbf{0}.$$

4.1.2.2. ANCOVA: OLS using FWL Theorem

By the zero conditional mean assumption,

$$\mathrm{Cov}(\varepsilon_T, X) = 0, \mathrm{Cov}(\varepsilon_T, \varepsilon_Y) = 0.$$

1) Residualization

Let $\gamma_{\mathrm{ANCOVA,1B}}$ denote the ANCOVA estimator under Scenario 1, Case B. By the FWL theorem or regression anatomy, the ANCOVA estimator is defined as

$$\gamma_{ANCOVA,1B} = \frac{\mathrm{Cov}(Y, \tilde{T})}{\mathrm{Var}(\tilde{T})},$$

where $\tilde{T}$ denotes the residual of T after partially out X:

$$\tilde{T} = T - \frac{\mathrm{Cov}(T, X)}{\mathrm{Var}(X)} X.$$

2) Compute $\tilde{T}$

From the $T = \delta_T X + \varepsilon_T$,

$$\mathrm{Cov}(T, X) = \delta_T \mathrm{Var}(X) + \mathrm{Cov}(\varepsilon_T, X) = \delta_T \mathrm{Var}(X).$$

So

$$\tilde{T} = T - \delta_T X = \varepsilon_T.$$

Thus, the residual $\tilde{T}$ is equal to ε_T under this Scenario 1, Case B.

Hence, by the SCM

$$\mathrm{Cov}(X, \tilde{T}) = \mathrm{Cov}(X, \varepsilon_T) = 0,$$

$$\mathrm{Var}(\tilde{T}) = \mathrm{Var}(\varepsilon_T).$$

3) Compute numerator: $\mathrm{Cov}(Y, \tilde{T})$

$$\begin{aligned}
\mathrm{Cov}(Y, \tilde{T}) &= \mathrm{Cov}(\tau X + \gamma T + \varepsilon_Y, \varepsilon_T) \\
&= \tau \mathrm{Cov}(X, \varepsilon_T) + \gamma \mathrm{Cov}(T, \varepsilon_T) + \mathrm{Cov}(\varepsilon_Y, \varepsilon_T) \\
&= \gamma \mathrm{Cov}(T, \varepsilon_T) \\
&= \gamma \mathrm{Cov}(\delta_T X + \varepsilon_T, \varepsilon_T) \\
&= \gamma \{\delta_T \mathrm{Cov}(X, \varepsilon_T) + \mathrm{Var}(\varepsilon_T)\} \\
&= \gamma \mathrm{Var}(\varepsilon_T).
\end{aligned}$$

Thus, the direct effect of X on Y (τ) drops out because $\tilde{T}$ is orthogonal to X $(= \varepsilon_T)$.

4) Compute denominator: $\mathrm{Var}(\tilde{T})$

$$\mathrm{Var}(\tilde{T}) = \mathrm{Var}(\varepsilon_T).$$

5) Accordingly,

$$\boldsymbol{\gamma_{ANCOVA,1B}} = \frac{\mathrm{Cov}(Y, \tilde{T})}{\mathrm{Var}(\tilde{T})} = \gamma \cdot \frac{\mathrm{Var}(\varepsilon_T)}{\mathrm{Var}(\varepsilon_T)} = \gamma,$$

which means $\boldsymbol{Bias(\gamma_{ANCOVA,1B})_{OLS} = 0}$.

Therefore, $Bias(\gamma_{ANCOVA,1B})_G$ is identical to $Bias(\gamma_{ANCOVA,1B})_{OLS}$ which is still zero . Scenario 1, Case B differs from Case A only by adding a direct carryover effect $X \to Y$ with coefficient τ. This edge creates a potential confounding structure for the treatment effect because X now affects both T and Y, opening the backdoor path $T \leftarrow X \to Y$. Unlike Case A, this is precisely the setting where ANCOVA's conditioning step matters: by conditioning on X,

ANCOVA blocks $T \leftarrow X \rightarrow Y$, leaving only the causal path $T \rightarrow Y$, and the resulting structural bias remains zero. The FWL decomposition mirrors this interpretation: residualizing T on X again produces $\tilde{T} = \varepsilon_T$, and residualizing Y on X removes the entire X-driven component of Y, including the direct τX term (and the part of Y induced by X through T). Regressing the residualized outcome on $\tilde{T}$ therefore recovers γ. Thus, in both Case A and Case B, the graphical approach and the OLS (FWL) agree that ANCOVA has zero structural bias, with the key difference that in Case B the zero-bias result depends on blocking the backdoor path through X, whereas in Case A there is no such path to begin with.

4.1.2.3. Change Score Analysis: Graphical Approach

Under Scenario 1, Case B, there are three potential backdoor paths from *T* to *C*:

$$T \leftarrow X \rightarrow C,$$

$$T \leftarrow X \rightarrow Y \rightarrow C,$$

$$T \rightarrow Y \leftarrow X \rightarrow C.$$

The first and second backdoor paths are naturally open since they do not contain any collider, then the noncausal associations transmitted along these paths are given by their structural path coefficients:

1) $\boldsymbol{T \leftarrow X \rightarrow C : \delta_T \times (-1) = -\delta_T,}$

2) $\boldsymbol{T \leftarrow X \rightarrow Y \rightarrow C : \delta_T \times \tau \times (+1) = \tau\, \delta_T.}$

By contrast, the last path is not a backdoor path. According to Pearl's backdoor criterion, a path must satisfy the first condition to be considered a backdoor path, it must begin with an arrow into the treatment *T*. Only after this condition is satisfied can we evaluate if unconditioned colliders (i.e., nodes at which two arrowheads converge) along the path block the spurious association. Accordingly, regardless of *Y* being a collider, this is not considered a backdoor path. Thus, the

total structural bias under Scenario 1, Case B is the sum of the associations via open backdoor paths 1) and 2):

$$\boldsymbol{Bias(\gamma_{Change,1B})_G} = -\delta_T + \tau\,\delta_T = \boldsymbol{\delta_T\,(\tau - 1)}.$$

4.1.2.4. Change Score Analysis: OLS

Under Scenario 1, Case B, the covariances are given by:

$$Cov(X,T) = Cov(X, \delta_T X + \varepsilon_T) = \delta_T\, Var(X) + Cov(X, \varepsilon_T) = \delta_T \cdot 1 + 0 = \delta_T,$$

$$Cov(Y,T) = Cov(\tau X + \gamma T + \varepsilon_Y, T) = \tau\, Cov(X,T) + \gamma\, Var(T) + Cov(\varepsilon_Y, T)$$

$$= \tau\,\delta_T + \gamma \cdot 1 + 0 = \tau\,\delta_T + \gamma.$$

Let $\gamma_{\text{Change, 1B}}$ denotes the change score estimator under Scenario 1, Case B. Then,

$$\boldsymbol{\gamma_{Change,1B}} = Cov(Y,T) - Cov(X,T) = \tau\,\delta_T + \gamma - \delta_T = \gamma + \boldsymbol{\delta_T(\tau - 1)}$$

which means $\boldsymbol{Bias(\gamma_{Change,1B})_{OLS} = \delta_T(\tau - 1)}$.

Therefore, $Bias(\gamma_{Change,1B})_G$ is identical to $Bias(\gamma_{Change,1B})_{OLS}$. One of two conditions must be met to obtain the unbiased treatment effect: either the causal effect of the pretest *X* on the treatment T must be zero (i.e., $\delta_T = 0$), or the causal effect of the pretest *X* on the posttest *Y* equals one (i.e., $\tau = 1$), which means that the X is equal to Y. If $\tau < 1$, then $\delta_T > 0$ yields a negative bias, meaning the true treatment effect is underestimated.

4.2. Scenario 2: An Unobserved Confounder

4.2.1. Case A: When the pretest does not affect the posttest score

4.2.1.1. ANCOVA: Graphical Approach

The ANCOVA analytic model conditions only on pretest *X*. Thus, the adjustment set implied by the analytic model is $\{X\}$. Although *A* appears in the data-generating model, it is omitted from the analytic model and therefore is not conditioned on. In the corresponding DAG,

A is a common cause of both T and Y, and X is a pre-treatment variable caused by A that also caused T. Consequently, there are two potential backdoor paths from T to Y:

1) $T \longleftarrow X \longleftarrow A \longrightarrow Y$,

2) $\boldsymbol{T \leftarrow A \rightarrow Y}$**.**

Conditioning on X blocks path 1) because X is a non-collider on that path. However, conditioning on X does not block path 2), since A is not conditioned on. Therefore, $\{X\}$ does not satisfy the backdoor criterion for identifying the causal effect of T on Y, and the ANCOVA coefficient on T is generally biased for γ. Under this linear structural model, the remaining open backdoor path contributes a structural bias equal to the product of its path coefficients:

$$\boldsymbol{Bias(\gamma_{ANCOVA,2A})_G = \alpha_T\, \alpha_2}.$$

4.2.1.2. ANCOVA: OLS using FWL Theorem

By the zero conditional mean assumption,

$$\mathrm{Cov}(\varepsilon_T, X) = \mathrm{Cov}(\varepsilon_T, A) = \mathrm{Cov}(\varepsilon_Y, T) = \mathrm{Cov}(\varepsilon_X, A) = 0.$$

1) Define residuals $\tilde{T}$ and $\tilde{A}$.

By the regression anatomy (FWL theorem), the partial regression coefficient on T in the ANCOVA analytic model is

$$\gamma_{ANCOVA,2A} = \frac{\mathrm{Cov}(Y, \tilde{T})}{\mathrm{Var}(\tilde{T})},$$

where $\tilde{T}$ denotes the residual of T after partialing out X:

$$\tilde{T} = T - \frac{\mathrm{Cov}(T, X)}{\mathrm{Var}(X)} X.$$

In addition, define $\tilde{A}$ as the residual of A after partialing out X:

$$\tilde{A} = A - \frac{\mathrm{Cov}(A, X)}{\mathrm{Var}(X)} X.$$

(Proof) Let $\hat{A} = aX$ be the fitted value from projecting A on X in the population. Here we work with mean centered variables, $E(A) = E(X) = 0$, so $\hat{A} = aX$ without an intercept. Define the residual $\tilde{A} = A - \hat{A} = A - aX$ and choose a so that $\mathrm{Cov}(\tilde{A}, X) = 0$:

$$\mathrm{Cov}(A - aX, X) = \mathrm{Cov}(A, X) - a\mathrm{Var}(X) = 0 \quad \Rightarrow \quad a = \frac{\mathrm{Cov}(A, X)}{\mathrm{Var}(X)}.$$

The FWL theorem is used in the standard way to express the population-level ANCOVA estimator as $\gamma_{\mathrm{ANCOVA}} = \frac{\mathrm{Cov}(Y,\tilde{T})}{\mathrm{Var}(\tilde{T})}$, where $\tilde{T} = T - \frac{\mathrm{Cov}(T,X)}{\mathrm{Var}(X)}X$ is the residualized treatment after projecting T on X. The additional residual $\tilde{A}$ is introduced only as an algebraic device to rewrite the assumed data-generating model in the same residualized space induced by conditioning on X. Because residualization is a linear projection rather than a causal operation, it is always legitimate to decompose any variable into the part predictable from X and the remaining part orthogonal to X, even when the causal direction runs from A to X. This makes clear that any remaining confounding after conditioning on X must act through the component of the unobserved confounder not captured by X, which is exactly what $\widetilde{A}$ represents.

2) Express $\tilde{T}$ in terms of $\tilde{A}$.

From $T = \alpha_T A + \delta_T X + \varepsilon_T$,

$$\mathrm{Cov}(T, X) = \mathrm{Cov}(\delta_T X + \alpha_T A + \varepsilon_T, X) = \delta_T \mathrm{Var}(X) + \alpha_T \mathrm{Cov}(A, X).$$

Hence,

$$\tilde{T} = T - \left(\delta_T + \alpha_T \frac{\mathrm{Cov}(A, X)}{\mathrm{Var}(X)}\right) X = \alpha_T \left(A - \frac{\mathrm{Cov}(A, X)}{\mathrm{Var}(X)} X\right) + \varepsilon_T = \alpha_T \tilde{A} + \varepsilon_T.$$

3) Decompose the numerator $\mathrm{Cov}(Y, \tilde{T})$.

From $Y = \alpha_2 A + \gamma T + \varepsilon_Y$,

$$\mathrm{Cov}(Y, \tilde{T}) = \mathrm{Cov}(\alpha_2 A + \gamma T + \varepsilon_Y, \tilde{T}) = \gamma\, \mathrm{Cov}(T, \tilde{T}) + \alpha_2\, \mathrm{Cov}(A, \tilde{T}).$$

Compute $\mathrm{Cov}(T, \tilde{T})$ (let $b = \mathrm{Cov}(T, X)/\mathrm{Var}(X)$):

$$\mathrm{Cov}(T, \tilde{T}) = \mathrm{Cov}(T, T - bX) = \mathrm{Var}(T) - b\,\mathrm{Cov}(T, X) = \mathrm{Var}(T) - \frac{\mathrm{Cov}(T, X)^2}{\mathrm{Var}(X)}.$$

Also,

$$\mathrm{Var}(\tilde{T}) = \mathrm{Var}(T - bX) = \mathrm{Var}(T) + b^2\mathrm{Var}(X) - 2b\mathrm{Cov}(T, X) = \mathrm{Var}(T) - \frac{\mathrm{Cov}(T, X)^2}{\mathrm{Var}(X)}.$$

Therefore,

$$\mathrm{Cov}(T, \tilde{T}) = \mathrm{Var}(\tilde{T}).$$

By the same argument,

$$\mathrm{Cov}(A, \tilde{A}) = \mathrm{Var}(\tilde{A}).$$

Next, compute $\mathrm{Cov}(A, \tilde{T})$:

$$\mathrm{Cov}(A, \tilde{T}) = \mathrm{Cov}(A, \alpha_T\tilde{A} + \varepsilon_T) = \alpha_T\mathrm{Cov}(A, \tilde{A}) = \alpha_T\mathrm{Var}(\tilde{A}).$$

Thus,

$$\mathrm{Cov}(Y, \tilde{T}) = \gamma\,\mathrm{Var}(\tilde{T}) + \alpha_2\alpha_T\,\mathrm{Var}(\tilde{A}).$$

4) Plug into the regression anatomy formula.

$$\boldsymbol{\gamma_{ANCOVA,2A}} = \gamma + \boldsymbol{\alpha_2\alpha_T}\frac{\mathbf{Var}(\tilde{\boldsymbol{A}})}{\mathbf{Var}(\tilde{\boldsymbol{T}})}.$$

5) Residual variances.

$$\mathrm{Var}(\tilde{A}) = \mathrm{Var}(A) - \frac{\mathrm{Cov}(A, X)^2}{\mathrm{Var}(X)}.$$

Since

$$\mathrm{Cov}(A, X) = \mathrm{Cov}(A, \alpha_1 A + \varepsilon_X) = \alpha_1\mathrm{Var}(A),$$

Thus,

$$\mathrm{Var}(\tilde{A}) = \mathrm{Var}(A) - \frac{\alpha_1^2\mathrm{Var}(A)^2}{\mathrm{Var}(X)}.$$

Also,

$$\mathrm{Var}(\tilde{T}) = \mathrm{Var}(T) - \frac{\mathrm{Cov}(T,X)^2}{\mathrm{Var}(X)}.$$

Since

$$\mathrm{Cov}(T,X) = \mathrm{Cov}(\alpha_T A + \delta_T X + \varepsilon_T, X) = \delta_T \mathrm{Var}(X) + \alpha_T \mathrm{Cov}(A,X),$$

Thus,

$$\mathrm{Var}(\tilde{T}) = \mathrm{Var}(T) - \frac{\left(\delta_T \mathrm{Var}(X) + \alpha_T \mathrm{Cov}(A,X)\right)^2}{\mathrm{Var}(X)}.$$

6) Assuming standardization ($\mathrm{Var}(A) = \mathrm{Var}(X) = \mathrm{Var}(T) = 1$), then

$$\boldsymbol{\gamma_{ANCOVA,2A}} = \gamma + \boldsymbol{\alpha_2 \alpha_T} \frac{\mathbf{1} - \boldsymbol{\alpha_1^2}}{\mathbf{1} - (\boldsymbol{\delta_T} + \boldsymbol{\alpha_T \alpha_1})^2}$$

where the terms added to γ consititue $\boldsymbol{Bias(\gamma_{ANCOVA,2A})_{OLS}}$.

7) Unbiasedness conditions.

Assuming $\mathrm{Var}(\tilde{T}) > 0$, $\gamma_{\mathrm{ANCOVA},2A} = \gamma$ if at least one of the following holds:

i) $\alpha_T = 0$ (no $A \to T$),

ii) $\alpha_2 = 0$ (no $A \to Y$),

iii) $\mathrm{Var}(\tilde{A}) = 0$, i.e., after partialing out X from A, no residual variance in A remains. This means the pretest X is a perfect proxy for A; in other words, X is effectively the same as A in this setting.

Therefore, $Bias(\gamma_{ANCOVA,2A})_G$ is not in general identical to $Bias(\gamma_{ANCOVA,2A})_{OLS}$. The graphical approach identifies the core confounding component implied by the remaining open backdoor structure after conditioning on X, whereas the OLS result using the FWL theorem expresses the total ANCOVA structural bias as that core component multiplied by a residualization-induced scaling term determined by the residual moments after partialling out X.

Since the $Bias(\gamma_{ANCOVA,2A})_{OLS}$ indicates the total ANCOVA structural bias under Scenario 2, Case A, the ANCOVA estimator equals the causal treatment effect, $\gamma_{ANCOVA,2A} = \gamma$, whenever at least one of three conditions holds assuming $Var\ (\tilde{T}) > 0$. First, if latent ability A does not affect treatment assignment ($\alpha_T = 0$), then the ability related confounding path through $A \rightarrow T$ is absent. Second, if A has no direct effect on the posttest outcome ($\alpha_2 = 0$), then Acannot induce bias in the treatment effect even if it predicts T. Third, even when A affects both T and Y, ANCOVA remains unbiased if the pretest X perfectly captures A in the sense that residualizing A on X leaves no remaining variation, $Var\ (\tilde{A}) = 0$; equivalently, X is a perfect proxy for A. From the graphical approach, the first two conditions are recovered directly because they correspond to blocking the backdoor path $T \leftarrow A \rightarrow Y$**.** However, the graphical approach does not, by itself, recover the third condition, because $Var\ (\tilde{A}) = 0$ is a strength or measurement condition (i.e., perfect proxying) rather than a purely structural feature of the DAG. In practice, this perfect proxy condition is often unrealistic. More precisely, the third condition is better viewed not as an independent requirement for unbiasedness, but as a determinant of the residualization induced scaling of the core confounding bias: in the regression anatomy, the confounding component is multiplied by the factor $Var\ (\tilde{A})/Var\ (\tilde{T})$. When $0 <$ Var $(\tilde{A}) <$ Var $(\tilde{T})$, the factor is less than 1 and attenuates the core bias, whereas Var $(\tilde{A}) >$ Var $(\tilde{T})$, the factor exceeds 1 and amplifies the core bias.

4.2.1.3. Change Score Analysis: Graphical Approach

With the inclusion of the unobserved confounder A, additional backdoor paths from T to the change score C become *d*-connected through A's effects on both the pretest X and the posttest Y. Because both X and A are unobserved in the change score analytic model, these paths cannot be blocked, therefore contribute to structural confounding bias. Under the linearity assumption and

the path-tracing rules, the induced noncausal association between $T$and C is obtained by summing the products of path coefficients along all open backdoor paths connecting $T$and C. Under Scenario 2, Case A, there are four such open backdoor paths:

1) $\boldsymbol{T \leftarrow X \rightarrow C}$: $\boldsymbol{-\delta_T}$,

2) $\boldsymbol{T \leftarrow A \rightarrow X \rightarrow C}$: $\boldsymbol{-\alpha_T\alpha_1}$,

3) $\boldsymbol{T \leftarrow A \rightarrow Y \rightarrow C}$: $\boldsymbol{\alpha_T\alpha_2}$,

4) $\boldsymbol{T \leftarrow X \leftarrow A \rightarrow Y \rightarrow C}$: $\boldsymbol{\delta_T\alpha_1\alpha_2}$.

Summing up these components yields the structural bias of the change score estimator under Scenario 2, Case A: $\boldsymbol{Bias(\gamma_{Change,2A})_G = \delta_T(\alpha_1\alpha_2 - 1) + \alpha_T(\alpha_2 - \alpha_1)}$.

4.2.1.4. Change Score Analysis: OLS

Under Scenario 2, Case A, the change score estimator is expressed as:

$\boldsymbol{\gamma_{Change,2A}} = Cov(Y,T) - Cov(X,T) = \gamma + \alpha_2(\delta_T\alpha_1 + \alpha_T) - (\delta_T + \alpha_T\alpha_1) = \gamma +$ $\boldsymbol{\delta_T(\alpha_1\alpha_2 - 1) + \alpha_T(\alpha_2 - \alpha_1)}$, where the terms added to γ constitute $\boldsymbol{Bias(\gamma_{Change,2A})_{OLS}}$. Given the required covariance terms are derived as follows:

$$\mathrm{Cov}(X,T) = \mathrm{Cov}(\alpha_1 A + \varepsilon_X,\ \alpha_T A + \delta_T X + \varepsilon_T)$$

$= \delta_T \mathrm{Cov}(X,X) + \alpha_T \mathrm{Cov}(X,A) + \mathrm{Cov}(X,\varepsilon_T)$

$= \delta_T \mathrm{Var}(X) + \alpha_T\alpha_1 \mathrm{Var}(A) + 0$

$= \delta_T \cdot 1 + \alpha_T\alpha_1 \cdot 1 = \delta_T + \alpha_T\alpha_1.$

$$\mathrm{Cov}(Y,T) = \mathrm{Cov}(\alpha_2 A + \gamma T + \varepsilon_Y,\ T)$$

$= \gamma \mathrm{Cov}(T,T) + \alpha_2 \mathrm{Cov}(A,T) + \mathrm{Cov}(\varepsilon_Y,T)$

$= \gamma \mathrm{Var}(T) + \alpha_2 \mathrm{Cov}(A, \alpha_T A + \delta_T X + \varepsilon_T) + 0$

$= \gamma \cdot 1 + \alpha_2 [\delta_T \mathrm{Cov}(A,X) + \alpha_T \mathrm{Var}(A)]$

$= \gamma + \alpha_2(\delta_T\alpha_1 + \alpha_T).$

Therefore, with standardized variables, $Bias(\gamma_{Change,2A})_G$ is identical to $Bias(\gamma_{Change,2A})_{OLS}$. This exact agreement occurs because the change score estimator is a simple regression coefficient with no pretest residualization step, so the noncausal covariance between T and the change score C obtained by summing open-backdoor path-tracing products is the same quantity that enters the OLS regression-coefficient formula, unlike ANCOVA where partialling out X introduces an additional residual-variance scaling.

It follows that setting bias to zero gives unbiasedness conditions. A practically interpretable sufficient set begins with $\delta_T = 0$, meaning there is no causal influence of the pretest X on treatment assignment T (i.e., no dynamic causal relationship from the prior outcome to treatment). Under $\delta_T = 0$, the change score estimator is unbiased if either $\alpha_2 = \alpha_1$, which corresponds to the common trend (time invariant confounding) assumption that the total effect of the unobserved confounder A on the posttest Yunder control equals its effect on the pretest X, or $\alpha_T = 0$, meaning A has no direct causal effect on treatment assignment and thus does not confound T. In addition, unbiasedness could also occur through the bias-offsetting mechanism that the two biasing terms technically can offset each other in change score analysis (Kim & Steiner, 2021); $\delta_T(\alpha_1\alpha_2 - 1) + \alpha_T(\alpha_2 - \alpha_1) = 0,\ thus\ \delta_T(\alpha_1\alpha_2 - 1) = -\alpha_T(\alpha_2 - \alpha_1)$. However, as Lüdtke & Robitzsch (2025) pointed out, such an exact offsetting of different biasing terms is implausible whenever biasing terms are different from zero.

4.2.2. Case B: When the pretest affects the posttest

4.2.2.1. ANCOVA: Graphical Approach

Under Scenario 2, Case B, the assumed data-generating DAG includes the causal effect of the pretest on the posttest ($X \rightarrow Y$). To identify structural bias using the graphical approach, we enumerate potential backdoor paths from T to Y, that is, paths that begin with an arrow into T.

Because both X and A are causes of T in this DAG, the following backdoor paths connect T and Y:

1) $T \leftarrow X \rightarrow Y$,

2) $T \leftarrow X \leftarrow A \rightarrow Y$,

3) $\boldsymbol{T \leftarrow A \rightarrow Y}$,

4) $T \leftarrow A \rightarrow X \rightarrow Y$.

Conditioning on X blocks paths 1), 2), and 4) because Xlies on each of these paths as a non-collider and is included in the conditioning set. In particular, compared to Case A, Case B introduces the additional backdoor path $T \leftarrow X \rightarrow Y$ due to the direct effect of the pretest on the posttest ($X \rightarrow Y$), but this path is explicitly blocked by ANCOVA because ANCOVA conditions on X. However, conditioning on X does not block path 3) $T \leftarrow A \rightarrow Y$, because A is not conditioned on. Since A is unobserved and thus omitted from the ANCOVA analytic model, the confounding path $T \leftarrow A \rightarrow Y$ remains open, implying that $\{X\}$ does not satisfy the backdoor criterion for identifying the causal effect of T on Y. Therefore, the ANCOVA estimator is systematically biased for the causal effect of T on Y, and the structural bias is expressed as:

$$\boldsymbol{Bias(\gamma_{ANCOVA,2B})_G = \alpha_T\, \alpha_2}.$$

4.2.2.2. ANCOVA: OLS using FWL Theorem

By the zero conditional mean assumption,

$$\mathrm{Cov}(\varepsilon_T, X) = \mathrm{Cov}(\varepsilon_T, A) = \mathrm{Cov}(\varepsilon_Y, T) = \mathrm{Cov}(\varepsilon_X, A) = 0.$$

1) Define residuals $\tilde{T}$ and $\tilde{A}$.

By the regression anatomy (FWL) theorem, the partial regression coefficient on T in the ANCOVA analytic model is

$$\gamma_{ANCOVA,2B} = \frac{\mathrm{Cov}(Y, \tilde{T})}{\mathrm{Var}(\tilde{T})},$$

where $\tilde{T}$ denotes the residual of T after partialing out X:

$$\tilde{T} = T - \frac{\mathrm{Cov}(T, X)}{\mathrm{Var}(X)} X.$$

In addition, define $\tilde{A}$ as the residual of A after partialing out X:

$$\tilde{A} = A - \frac{\mathrm{Cov}(A, X)}{\mathrm{Var}(X)} X.$$

2) Express $\tilde{T}$ in terms of $\tilde{A}$.

From $T = \alpha_T A + \delta_T X + \varepsilon_T$,

$$\mathrm{Cov}(T, X) = \mathrm{Cov}(\alpha_T A + \delta_T X + \varepsilon_T, X) = \delta_T \mathrm{Var}(X) + \alpha_T \mathrm{Cov}(A, X).$$

Hence,

$$\tilde{T} = T - \left(\delta_T + \alpha_T \frac{\mathrm{Cov}(A, X)}{\mathrm{Var}(X)}\right) X = \alpha_T \left(A - \frac{\mathrm{Cov}(A, X)}{\mathrm{Var}(X)} X\right) + \varepsilon_T = \alpha_T \tilde{A} + \varepsilon_T.$$

3) Decompose the numerator $\mathrm{Cov}(Y, \tilde{T})$.

From $Y = \alpha_2 A + \tau X + \gamma T + \varepsilon_Y$,

$$\mathrm{Cov}(Y, \tilde{T}) = \mathrm{Cov}(\alpha_2 A + \tau X + \gamma T + \varepsilon_Y, \tilde{T}).$$

Thus,

$$\mathrm{Cov}(Y, \tilde{T}) = \tau \mathrm{Cov}(X, \tilde{T}) + \gamma \mathrm{Cov}(T, \tilde{T}) + \alpha_2 \mathrm{Cov}(A, \tilde{T}) + \mathrm{Cov}(\varepsilon_Y, \tilde{T}).$$

Because $\mathrm{Cov}(X, \tilde{T}) = 0$ and $\mathrm{Cov}(\varepsilon_Y, \tilde{T}) = 0$,

$$\mathrm{Cov}(Y, \tilde{T}) = \gamma \mathrm{Cov}(T, \tilde{T}) + \alpha_2 \mathrm{Cov}(A, \tilde{T}).$$

Also,

$$\mathrm{Cov}(T, \tilde{T}) = \mathrm{Var}(\tilde{T}),$$

$$\mathrm{Cov}(A, \tilde{A}) = \mathrm{Var}(\tilde{A}).$$

Next, compute $\mathrm{Cov}(A, \tilde{T})$:

$$\mathrm{Cov}(A, \tilde{T}) = \mathrm{Cov}(A, \alpha_T \tilde{A} + \varepsilon_T) = \alpha_T \mathrm{Cov}(A, \tilde{A}) = \alpha_T \mathrm{Var}(\tilde{A}).$$

Accordingly,

$$\mathrm{Cov}(Y, \tilde{T}) = \gamma \mathrm{Var}(\tilde{T}) + \alpha_2 \alpha_T \mathrm{Var}(\tilde{A}).$$

4) Plug into the regression anatomy formula.

$$\boldsymbol{\gamma_{ANCOVA,2B}} = \gamma + \boldsymbol{\alpha_2 \alpha_T} \frac{\boldsymbol{Var}(\tilde{\boldsymbol{A}})}{\boldsymbol{Var}(\tilde{\boldsymbol{T}})}.$$

5) Residual variances.

$$\mathrm{Var}(\tilde{A}) = \mathrm{Var}(A) - \frac{\mathrm{Cov}(A, X)^2}{\mathrm{Var}(X)}.$$

Since

$$\mathrm{Cov}(A, X) = \mathrm{Cov}(A, \alpha_1 A + \varepsilon_X) = \alpha_1 \mathrm{Var}(A),$$

Thus,

$$\mathrm{Var}(\tilde{A}) = \mathrm{Var}(A) - \frac{\alpha_1^2 \mathrm{Var}(A)^2}{\mathrm{Var}(X)}.$$

Also,

$$\mathrm{Var}(\tilde{T}) = \mathrm{Var}(T) - \frac{\mathrm{Cov}(T, X)^2}{\mathrm{Var}(X)}.$$

Since

$$\mathrm{Cov}(T, X) = \mathrm{Cov}(\alpha_T A + \delta_T X + \varepsilon_T, X) = \delta_T \mathrm{Var}(X) + \alpha_T \mathrm{Cov}(A, X),$$

Thus,

$$\mathrm{Var}(\tilde{T}) = \mathrm{Var}(T) - \frac{\left(\delta_T \mathrm{Var}(X) + \alpha_T \mathrm{Cov}(A, X)\right)^2}{\mathrm{Var}(X)}.$$

6) Assuming standardization ($\mathrm{Var}(A) = \mathrm{Var}(X) = \mathrm{Var}(T) = 1$), then

$$\boldsymbol{\gamma_{ANCOVA,2B}} = \gamma + \boldsymbol{\alpha_2\alpha_T}\frac{\mathbf{1} - \boldsymbol{\alpha_1^2}}{\mathbf{1} - (\boldsymbol{\delta_T} + \boldsymbol{\alpha_T\alpha_1})^2}$$

where the terms added to γ constitute $\boldsymbol{Bias(\gamma_{ANCOVA,2B})_{OLS}}$, which is the same as $\boldsymbol{Bias(\gamma_{ANCOVA,2A})_{OLS}}$.

Under Scenario 2, Case B, the data generating DAG extends Case A by adding a direct pretest to posttest effect, $X \rightarrow Y$, with coefficient τ. Although this additional edge opens a new potential backdoor path from T to Y through X (specifically, $T \leftarrow X \rightarrow Y$), it does not change the ANCOVA structural bias relative to Case A because ANCOVA conditions on X. Conditioning on X blocks any backdoor path that runs through X, so the only backdoor path that remains open is the same core confounding path $T \leftarrow A \rightarrow Y$ induced by the unobserved ability confounder A. As a result, the graphical approach again identifies the same core noncausal confounding bias as in Case A and yields the same sufficient unbiasedness conditions, because those conditions correspond to eliminating one of the two edges of the remaining open backdoor path (no $A \rightarrow T$ or no $A \rightarrow Y$). The regression anatomy confirms this invariance to adding $X \rightarrow Y$. Under the FWL identity, the ANCOVA coefficient can be written as $\gamma_{ANCOVA} = \mathrm{Cov}\,(Y, \tilde{T})/\mathrm{Var}\,(\tilde{T})$, where $\tilde{T}$ is the residual from regressing T on X. Because $\mathrm{Cov}\,(X, \tilde{T}) = 0$ by construction, the direct contribution of $X$to Y via τXcannot affect $\mathrm{Cov}\,(Y, \tilde{T})$and therefore does not enter the ANCOVA bias. Thus, in Case B, the graphical approach and the OLS agree on the unbiasedness conditions and on the fact that the only remaining source of bias is the core confounding component through A. To quantify the exact structural bias, it suffices to take the core confounding bias identified by the graphical approach and multiply it by the residualization induced variance scaling factor from the regression anatomy.

4.2.2.3. Change Score Analysis: Graphical Approach

Under Scenario 2, Case B, the additional direct effect of the pretest on the posttest ($X \rightarrow Y$ with a path coefficient τ) creates further backdoor paths from T to C that run through X and Y. Compared to Case A, this direct edge adds two additional backdoor paths, so there are now six open backdoor paths from T to C.

1) $\boldsymbol{T \leftarrow X \rightarrow C}$: $\boldsymbol{-\delta_T}$,
2) $\boldsymbol{T \leftarrow X \rightarrow Y \rightarrow C}$: $\boldsymbol{\delta_T \tau}$,
3) $\boldsymbol{T \leftarrow A \rightarrow X \rightarrow C}$: $\boldsymbol{-\alpha_T \alpha_1}$,
4) $\boldsymbol{T \leftarrow A \rightarrow Y \rightarrow C}$: $\boldsymbol{\alpha_T \alpha_2}$,
5) $\boldsymbol{T \leftarrow A \rightarrow X \rightarrow Y \rightarrow C}$: $\boldsymbol{\alpha_T \alpha_1 \tau}$,
6) $\boldsymbol{T \leftarrow X \leftarrow A \rightarrow Y \rightarrow C}$: $\boldsymbol{\delta_T \alpha_1 \alpha_2}$.

Summing up these products constitutes the structural bias of the change score estimator under Scenario 2, Case B: $\boldsymbol{Bias(\gamma_{Change,2B})_G = \delta_T(\tau + \alpha_1\alpha_2 - 1) + \alpha_T[\alpha_1(\tau - 1) + \alpha_2]}$.

4.2.2.4. Change Score Analysis: OLS

Under Scenario 2, Case B, the change score estimator is expressed as:

$$\boldsymbol{\gamma_{Change,\ 2B}} = Cov(Y,T) - Cov(X,T)$$

$= \gamma + \boldsymbol{\delta_T(\tau + \alpha_1\alpha_2 - 1) + \alpha_T[\alpha_1(\tau - 1) + \alpha_2]}$ where the terms added to γ constitute $\boldsymbol{Bias(\gamma_{Change,2B})_{OLS}}$. Given the required covariance terms are derived as follows:

$\mathrm{Cov}(X,T) = \mathrm{Cov}(\alpha_1 A + \varepsilon_X,\ \alpha_T A + \delta_T X + \varepsilon_T) = \delta_T \mathrm{Var}(X) + \alpha_T\alpha_1 \mathrm{Var}(A) = \delta_T + \alpha_T\alpha_1,$

$\mathrm{Cov}(Y,T) = \mathrm{Cov}(\alpha_2 A + \tau X + \gamma T + \varepsilon_Y,\ T) = \tau\,\mathrm{Cov}(X,T) + \gamma\,\mathrm{Var}(T) + \alpha_2\,\mathrm{Cov}(A,T)$

$= \tau\,(\delta_T + \alpha_T\alpha_1) + \gamma + \alpha_2(\delta_T\alpha_1 + \alpha_T).$

Therefore, with standardized variables, $Bias(\gamma_{Change,2B})_G$ is identical to $Bias(\gamma_{Change,2B})_{OLS}$. Again, This equality holds because the change-score estimator is a simple

regression coefficient, $\gamma_{\text{Change}} = \text{Cov}(C, T)/\text{Var}(T)$, and the path-tracing sum over open backdoor paths gives exactly $\text{Cov}(C, T)$ (with the same $\text{Var}(T)$ denominator), so no additional residualization-based scaling is introduced. Under Scenario 2, Case B, the data generating DAG extends Case A by allowing a direct carryover from the pretest to the posttest, $X \rightarrow Y$, with coefficient τ. Despite this added causal link, the change score estimator remains a simple regression of Con T, so under variable standardization the graphical approach and the OLS yield the same structural bias expression and therefore imply the same unbiasedness conditions.

Accordingly, unbiasedness can be stated in an interpretable way as follows. First, if $\delta_T = 0$, meaning there is no dynamic causal relationship from the pretest to treatment assignment (no $X \rightarrow T$), then unbiasedness additionally requires the common trend (time-invariant confounding) condition $\alpha_1(\tau - 1) + \alpha_2 = 0$, equivalently $\alpha_2 + \tau\alpha_1 = \alpha_1$. This condition says that the total effect of the unobserved confounder Aon the untreated posttest Y, combining its direct effect α_2 and its indirect effect transmitted through Xvia $\tau\alpha_1$, matches its effect on the pretest X. Notably, when $\tau = 0$, this reduces to the familiar $\alpha_2 = \alpha_1$ condition from Case A, whereas when $\tau = 1$, it reduces to $\alpha_2 = 0$, meaning that the pretest differences carry forward perfectly to the posttest. Alternatively, unbiasedness also holds if $\alpha_T = 0$, so that A has no causal effect on treatment assignment and cannot confound T. In addition, unbiasedness is also technically possible through exact bias offsetting, which occurs when $\delta_T(\tau + \alpha_1\alpha_2 - 1) = -\alpha_T[\alpha_1(\tau - 1) + \alpha_2]$.

4.2.3. Case C: When the pretest affects the posttest through a mediator

4.2.3.1. ANCOVA: Graphical Approach

Under Scenario 2, Case C, the following potential backdoor paths are blocked because ANCOVA conditions on X:

1) $T \leftarrow X \rightarrow Y$,

2) $T \leftarrow X \leftarrow A \rightarrow Y$,

3) $T \leftarrow X \leftarrow A \rightarrow M \rightarrow Y$,

4) $T \leftarrow A \rightarrow X \rightarrow Y$,

5) $T \leftarrow A \rightarrow X \rightarrow M \rightarrow Y$.

Although these two following paths connect $T$and Y, it does not qualify as backdoor paths because it starts with an arrow out of T, not into it. If we were to condition on M, collider Mwould open the path and create a noncausal association between $T$and Y; however, M is not conditioned on in ANCOVA, so the unobserved collider M blocks the paths, and no bias is transmitted through it:

1) $T \rightarrow M \leftarrow A \rightarrow Y$,

2) $T \rightarrow M \leftarrow A \rightarrow X \rightarrow Y$.

Then, there are two open backdoor paths from T to Y:

1) $\boldsymbol{T \leftarrow A \rightarrow Y}$**:** $\boldsymbol{\alpha_T \alpha_2}$**,**

2) $\boldsymbol{T \leftarrow A \rightarrow M \rightarrow Y}$**:** $\boldsymbol{\alpha_T \alpha_M \eta}$**.**

In addition, this case involves a conceptually distinct source of bias that is not due to an open backdoor path. Because the treatment effect is defined as the direct effect of Ton Y(i.e., the $T \rightarrow Y$ edge), omitting the mediator M from the analytic model causes the ANCOVA estimator to conflate the direct effect with the mediated effect. Consequently, the indirect causal chain $T \rightarrow M \rightarrow Y$is absorbed into the coefficient on T, yielding an omitted mediator bias equal to the product of the path coefficients:

3) $\boldsymbol{T \rightarrow M \rightarrow Y}$**:** $\boldsymbol{\zeta\eta}$**.**

Thus,

$$\boldsymbol{Bias}(\boldsymbol{\gamma_{ANCOVA,2C}})_G = \boldsymbol{\zeta\eta} + \boldsymbol{\alpha_T}(\boldsymbol{\eta\alpha_M} + \boldsymbol{\alpha_2}).$$

4.2.3.2. ANCOVA: OLS using FWL theorem

By the zero conditional mean assumption,

$\mathrm{Cov}(\varepsilon_T, X) = \mathrm{Cov}(\varepsilon_T, A) = 0,$

$\mathrm{Cov}(\varepsilon_M, X) = \mathrm{Cov}(\varepsilon_M, T) = \mathrm{Cov}(\varepsilon_M, A) = 0,$

$\mathrm{Cov}(\varepsilon_Y, X) = \mathrm{Cov}(\varepsilon_Y, T) = \mathrm{Cov}(\varepsilon_Y, M) = \mathrm{Cov}(\varepsilon_Y, A) = 0,$ and all error terms are mutually uncorrelated.

1) Define residuals $\tilde{T}$ and $\tilde{A}$

By the regression anatomy (FWL) theorem, the partial regression coefficient on T in the ANCOVA analytic model is

$$\gamma_{ANCOVA,2C} = \frac{\mathrm{Cov}(Y, \tilde{T})}{\mathrm{Var}(\tilde{T})}.$$

Here $\tilde{T}$ denotes the residual of T after partialling out X:

$$\tilde{T} = T - \frac{\mathrm{Cov}(T, X)}{\mathrm{Var}(X)} X.$$

In addition, define $\tilde{A}$ as the residual of A after partialling out X:

$$\tilde{A} = A - \frac{\mathrm{Cov}(A, X)}{\mathrm{Var}(X)} X.$$

2) Express $\tilde{T}$ in terms of $\tilde{A}$

Start from the treatment equation,

$$T = \alpha_T A + \delta_T X + \varepsilon_T.$$

First compute $\mathrm{Cov}(T, X)$:

$$\mathrm{Cov}(T, X) = \mathrm{Cov}(\alpha_T A + \delta_T X + \varepsilon_T, X)$$

$$= \delta_T \mathrm{Var}(X) + \alpha_T \mathrm{Cov}(A, X) + \mathrm{Cov}(\varepsilon_T, X)$$

$$= \delta_T \mathrm{Var}(X) + \alpha_T \mathrm{Cov}(A, X).$$

Plugging this into the residual definition,

$$\tilde{T} = T - \frac{\mathrm{Cov}(T, X)}{\mathrm{Var}(X)}\, X$$

$$= \delta_T X + \alpha_T A + \varepsilon_T - \frac{\delta_T \mathrm{Var}(X) + \alpha_T \mathrm{Cov}(A, X)}{\mathrm{Var}(X)}\, X$$

$$= \alpha_T \left(A - \frac{\mathrm{Cov}(A, X)}{\mathrm{Var}(X)}\, X \right) + \varepsilon_T$$

$$= \alpha_T \tilde{A} + \varepsilon_T.$$

3) Reduce Y to a function of X, T, and A

Substitute the mediator equation into the outcome equation:

$$M = \alpha_M A + \delta_M X + \zeta T + \varepsilon_M,$$

$$Y = \alpha_2 A + \tau X + \gamma T + \eta M + \varepsilon_Y.$$

Then

$$Y = \alpha_2 A + \tau X + \gamma T + \eta(\alpha_M A + \delta_M X + \zeta T + \varepsilon_M) + \varepsilon_Y$$

$$= (\alpha_2 + \eta\alpha_M) A + (\tau + \eta\delta_M) X + (\gamma + \eta\zeta) T + (\eta\varepsilon_M + \varepsilon_Y).$$

Define the shorthand

$$\tau_c = \tau + \eta\delta_M,$$

$$\gamma_c = \gamma + \eta\zeta,$$

$$\alpha_{2_c} = \alpha_2 + \eta\alpha_M,$$

$$\varepsilon_{Y_c} = \eta\varepsilon_M + \varepsilon_Y,$$

so that

$$Y = \tau_c X + \gamma_c T + \alpha_{2_c} A + \varepsilon_{Y_c}.$$

4) Decompose the numerator $\mathrm{Cov}\left(Y, \tilde{T}\right)$

Using the reduced form,

$$\mathrm{Cov}\big(Y,\tilde{T}\big)=\mathrm{Cov}\big(\alpha_{2_c}A+\tau_cX+\gamma_cT+\varepsilon_{Y_c},\tilde{T}\big)$$

$$=\alpha_{2_c}\mathrm{Cov}\big(A,\tilde{T}\big)+\tau_c\mathrm{Cov}\big(X,\tilde{T}\big)+\gamma_c\mathrm{Cov}\big(T,\tilde{T}\big)+\mathrm{Cov}\big(\varepsilon_{Y_c},\tilde{T}\big).$$

Since $\mathrm{Cov}\big(X,\tilde{T}\big)=0$, $\mathrm{Cov}\big(\varepsilon_{Y_c},\tilde{T}\big)=0$, $\mathrm{Cov}\big(T,\tilde{T}\big)=\mathrm{Var}\big(\tilde{T}\big)$, and

$$\mathrm{Cov}\big(A,\tilde{T}\big)=\mathrm{Cov}\big(A,\alpha_T\tilde{A}+\varepsilon_T\big).$$

$$=\alpha_T\mathrm{Cov}\big(A,\tilde{A}\big)+\mathrm{Cov}(A,\varepsilon_T).$$

$$=\alpha_T\mathrm{Var}\big(\tilde{A}\big).$$

Putting these together,

$$\mathrm{Cov}\big(Y,\tilde{T}\big)=\gamma_c\mathrm{Var}\big(\tilde{T}\big)+\alpha_{2_c}\alpha_T\mathrm{Var}\big(\tilde{A}\big).$$

5) Apply the decomposition to the regression anatomy formula

$$\boldsymbol{\gamma_{ANCOVA,2C}}=\frac{\mathrm{Cov}\big(Y,\tilde{T}\big)}{\mathrm{Var}\big(\tilde{T}\big)}$$

$$=\gamma_c+\alpha_{2_c}\alpha_T\frac{\mathrm{Var}\big(\tilde{A}\big)}{\mathrm{Var}\big(\tilde{T}\big)}$$

$$=\gamma+\boldsymbol{\eta\zeta}+(\boldsymbol{\alpha_2}+\boldsymbol{\eta\alpha_M})\boldsymbol{\alpha_T}\frac{\mathbf{Var}\big(\widetilde{\boldsymbol{A}}\big)}{\mathbf{Var}\big(\widetilde{\boldsymbol{T}}\big)}.$$

6) Residual variances

$$\mathrm{Var}\big(\tilde{A}\big)=\mathrm{Var}(A)-\frac{\mathrm{Cov}(A,X)^2}{\mathrm{Var}(X)},$$

$$\mathrm{Var}\big(\tilde{T}\big)=\mathrm{Var}(T)-\frac{\mathrm{Cov}(T,X)^2}{\mathrm{Var}(X)}.$$

From $X=\alpha_1A+\varepsilon_X$ and $\mathrm{Cov}(\varepsilon_X,A)=0$,

$$\mathrm{Cov}(A,X)=\mathrm{Cov}(A,\alpha_1A+\varepsilon_X)=\alpha_1\mathrm{Var}(A).$$

From $T=\alpha_TA+\delta_TX+\varepsilon_T$,

$$\mathrm{Cov}(T,X)=\mathrm{Cov}(\alpha_TA+\delta_TX+\varepsilon_T,X)$$

$$= \delta_T \mathrm{Var}(X) + \alpha_T \mathrm{Cov}(A, X)$$

$$= \delta_T \mathrm{Var}(X) + \alpha_T \alpha_1 \mathrm{Var}(A).$$

7) Assuming standardization

$$\mathrm{Var}(\tilde{A}) = 1 - \alpha_1^2,$$

$$\mathrm{Var}(\tilde{T}) = 1 - (\delta_T + \alpha_T \alpha_1)^2.$$

Thus,

$$\boldsymbol{\gamma_{ANCOVA,2C}} = \gamma + \boldsymbol{\eta\zeta} + (\boldsymbol{\alpha_2} + \boldsymbol{\eta\alpha_M})\boldsymbol{\alpha_T} \frac{\mathbf{1} - \boldsymbol{\alpha_1^2}}{\mathbf{1} - (\boldsymbol{\delta_T} + \boldsymbol{\alpha_T\alpha_1})^2}$$

where the terms added to γ constitute $\boldsymbol{Bias(\gamma_{ANCOVA,2C})_{OLS}}$.

8) Unbiasedness conditions

Assuming $\mathrm{Var}(\tilde{T}) > 0$, unbiasedness holds if

$$\eta\zeta = 0 \quad \text{(i.e., } \eta = 0 \text{ or } \zeta = 0\text{)},$$

and

$$(\alpha_2 + \eta\alpha_M)\alpha_T(1 - \alpha_1^2) = 0,$$

which is satisfied if at least one of the following holds:

i. $\alpha_T = 0$ (no $A \to T$),

ii. $\alpha_2 + \eta\alpha_M = 0$ (no $A \to Y$ *and* $A \to M \to Y$),

iii. $\mathrm{Var}(\tilde{A}) = 0$ (X as a perfect proxy for A).

Therefore, $Bias(\gamma_{ANCOVA,2C})_G$ captures the core confounding component identified by the graphical approach, whereas $Bias(\gamma_{ANCOVA,2C})_{OLS}$ expresses the total ANCOVA structural bias with the residualization-induced scaling implied by the FWL theorem. Under Scenario 2, Case C, the data generating DAG extends Case B by introducing an unobserved mediator M between treatment and the posttest, while retaining the direct pretest to posttest effect $X \to Y$. As in Case

B, adding new edges creates additional potential paths linking T and Y, but ANCOVA conditions on X, so any path that runs through X is blocked and does not contribute to the ANCOVA bias. The key difference from Cases A and B is that conditioning on X cannot eliminate paths that operate through the unobserved mediator M or through A's influence on M. In particular, two noncausal backdoor paths remain open after conditioning on X: the familiar confounding path $T \leftarrow A \rightarrow Y$ and the additional confounding path $T \leftarrow A \rightarrow M \rightarrow Y$. The graphical decomposition therefore identifies a core confounding component that now combines A's direct effect on Y with its indirect effect transmitted via M: $\alpha_{\mathrm{T}}(\alpha_2 + \eta\alpha_{\mathrm{M}})$. Case C also introduces a second, conceptually distinct source of structural bias that was absent in Cases A and B: because M is omitted from the ANCOVA analytic model while the target causal treatment effect is the direct effect γ, the causal mediation path $T \rightarrow M \rightarrow Y$ contributes an omitted mediator component (of magnitude $\eta\zeta$) to the difference between γ_{ANCOVA} and γ. The OLS regression anatomy confirms this structure: the mediator component enters additively, whereas the confounding component enters through the same residualization induced variance scaling logic as in Cases A and B. Thus, the graphical approach continues to recover the relevant unbiasedness conditions associated with blocking the remaining A based confounding (now expressed in terms of the total A → Y effect, $\alpha_2 + \eta\alpha_{\mathrm{M}}$), but unbiasedness in Case C additionally requires that the mediation path be absent (that is, $\eta\zeta = 0$). For exact numerical bias quantification, the core confounding bias identified by the graphical approach must be scaled by the residualization induced variance factor, while the omitted mediator component is added as the separate mediation term.

4.2.3.3. Change Score Analysis: Graphical Approach

Under Scenario 2, Case C, the additional indirect effect of the pretest on the posttest through the mediator M ($X \rightarrow M \rightarrow Y$) creates further backdoor paths from T to C that run through

X, M, and Y. Compared to Case B, this indirect effect adds two additional backdoor paths, so there are now eight open backdoor paths from T to C:

1) $\boldsymbol{T \leftarrow X \rightarrow C}$: $\boldsymbol{-\delta_T}$,
2) $\boldsymbol{T \leftarrow X \rightarrow M \rightarrow Y \rightarrow C}$: $\boldsymbol{\delta_T \delta_M \eta}$,
3) $\boldsymbol{T \leftarrow X \leftarrow A \rightarrow Y \rightarrow C}$: $\boldsymbol{\delta_T \alpha_1 \alpha_2}$,
4) $\boldsymbol{T \leftarrow X \leftarrow A \rightarrow M \rightarrow Y \rightarrow C}$: $\boldsymbol{\delta_T \alpha_1 \alpha_M \eta}$,
5) $\boldsymbol{T \leftarrow A \rightarrow X \rightarrow C}$: $\boldsymbol{-\alpha_T \alpha_1}$,
6) $\boldsymbol{T \leftarrow A \rightarrow Y \rightarrow C}$: $\boldsymbol{\alpha_T \alpha_2}$,
7) $\boldsymbol{T \leftarrow A \rightarrow X \rightarrow M \rightarrow Y \rightarrow C}$: $\boldsymbol{\alpha_T \alpha_1 \delta_M \eta}$,
8) $\boldsymbol{T \leftarrow A \rightarrow M \rightarrow Y \rightarrow C}$: $\boldsymbol{\alpha_T \alpha_M \eta}$.

In addition, because the treatment effect is defined solely as the direct effect of T on Y (i.e., the $T \rightarrow Y$ edge) while the mediator M is omitted from the change score analytic model, the mediated causal effect of T on Y is not separated from the direct effect. Instead, it is carried into the change score outcome C through the mapping $Y \rightarrow C$ (e.g., when $C = Y - X$, any change in Y enters C one-for-one). Thus, even though the following path is fully causal (not a spurious backdoor path), it contributes to structural bias as an omitted mediator bias:

$\boldsymbol{T \rightarrow M \rightarrow Y \rightarrow C}$: $\boldsymbol{\zeta \eta}$.

Summing up these products yields the structural bias of the change score estimator under Scenario 2, Case C:

$$\boldsymbol{Bias(\gamma_{Change,2C})_G}$$

$$= \boldsymbol{\eta\zeta + \delta_T(\eta\delta_M + \eta\alpha_1\alpha_M + \alpha_1\alpha_2 - 1) + \alpha_T(\eta\alpha_1\delta_M + \eta\alpha_M + \alpha_2 - \alpha_1)}.$$

4.2.3.4. Change Score Analysis: OLS

Under Scenario 2, Case C, the change score estimator is expressed as:

$$\boldsymbol{\gamma_{Change,\,2C}} = Cov(Y,T) - Cov(X,T)$$

$= \gamma + \boldsymbol{\eta\zeta + \delta_T(\eta\delta_M + \eta\alpha_1\alpha_M + \alpha_1\alpha_2 - 1) + \alpha_T(\eta\alpha_1\delta_M + \eta\alpha_M + \alpha_2 - \alpha_1)}$ where the terms added to γ constitute $\boldsymbol{Bias(\gamma_{Change,2C})_{OLS}}$. Given the required covariance terms are derived as follows:

$$\mathrm{Cov}(X,T) = \delta_T + \alpha_T\alpha_1,$$

$$\mathrm{Cov}(Y,T) = \gamma + \eta[\delta_M(\delta_T + \alpha_T\alpha_1) + \zeta + \alpha_M(\delta_T\alpha_1 + \alpha_T)] + \alpha_2(\delta_T\alpha_1 + \alpha_T).$$

Therefore, $Bias(\gamma_{Change,2C})_G$ remains identical to $Bias(\gamma_{Change,2C})_{OLS}$. Under Scenario 2, Case C, the data generating DAG extends Case B by introducing an unobserved mediator M that links the pretest and treatment to the posttest (via $X \rightarrow M \rightarrow Y$and $T \rightarrow M \rightarrow Y$). Relative to Cases A and B, this added mediation structure opens additional backdoor paths from $T$to the change score outcome C that run through X, M, and Y. At the same time, because the treatment effect of interest is defined as the direct effect of Ton Y while the change score analytic model does not include M, the fully causal mediated path $T \rightarrow M \rightarrow Y$contributes to structural bias as an omitted mediator component (carried into C through the mapping from Y to C). With standardized variables, the graphical approach and the OLS results again yield the same structural bias expression for the change score estimator and therefore imply the same unbiasedness conditions.

Accordingly, unbiasedness in Case C requires, first, that there be no causal mediation of the treatment effect through M, $\eta\zeta = 0$, so either Tdoes not affect $M(\zeta = 0)$or M does not affect $Y(\eta = 0)$. Second, the baseline driven treatment assignment mechanism must be absent, $\delta_T = 0$ (i.e., no dynamic causal relationship from X to T). Under $\delta_T = 0$, unbiasedness additionally requires the generalized common trend (i.e., time invariant confounding) condition $\eta\alpha_1\delta_M + \eta\alpha_M + \alpha_2 - \alpha_1 = 0$, equivalently $\alpha_1 = \alpha_2 + \eta(\alpha_M + \alpha_1\delta_M)$, which states that the total effect of the unobserved confounder Aon the untreated posttest Y, combining its direct effect on $Y$and its

indirect effects transmitted through M (including the $A \rightarrow X \rightarrow M \rightarrow Y$ component), matches its effect on the pretest X. Alternatively, unbiasedness also holds if $\alpha_T = 0$, so that A has no causal effect on treatment assignment and cannot confound T. In addition, unbiasedness is also technically possible through exact bias offsetting, which occurs when $\delta_T(\eta\delta_M + \eta\alpha_1\alpha_M + \alpha_1\alpha_2 - 1) = -\alpha_T(\eta\alpha_1\delta_M + \eta\alpha_M + \alpha_2 - \alpha_1)$.

4.3. Scenario 3: Two Unobserved Confounders

4.3.1. Case A: When the pretest does not affect the posttest

4.3.1.1. ANCOVA: Graphical Approach

Under Scenario 3, Case A, the data generating model extends the previous scenario by adding a second unobserved confounder B (in addition to A) and allowing A and B to be correlated through a common cause U. The ANCOVA analytic model conditions only on the pretest X (see Equation (1)), so the implied adjustment set remains $\{X\}$, while both A and B are omitted and therefore not conditioned on. To streamline the exposition from Scenario 3 onward, I focus only on the backdoor paths that remain open under the analytic adjustment set. For notational simplicity, let $\rho = \rho_A\rho_B$ under standardization.

Under the corresponding DAG, there are six open backdoor paths from T to Y. Two are the direct confounding paths through each unobserved confounder:

$$\mathbf{(1)}\ \boldsymbol{T \leftarrow A \rightarrow Y}\text{: } \boldsymbol{\alpha_T\alpha_2}, \qquad \mathbf{(2)}\ \boldsymbol{T \leftarrow B \rightarrow Y}\text{: } \boldsymbol{\beta_T\beta_2},$$

and two additional open paths arise from the correlation between A and B:

$$\mathbf{(3)}\ \boldsymbol{T \leftarrow A \leftarrow U \rightarrow B \rightarrow Y}\text{: } \boldsymbol{\alpha_T\rho\beta_2}, \qquad \mathbf{(4)}\ \boldsymbol{T \leftarrow B \leftarrow U \rightarrow A \rightarrow Y}\text{: } \boldsymbol{\beta_T\rho\alpha_2}.$$

Finally, because the pretest X is modeled as a common effect of A and B (that is, $X = \alpha_1 A + \beta_1 B + \varepsilon_X$), conditioning on X in ANCOVA opens the collider paths:

$$\mathbf{(5)}\ \boldsymbol{T \leftarrow A \rightarrow X \leftarrow B \rightarrow Y}\text{: } \boldsymbol{\alpha_T\alpha_1\beta_1\beta_2}, \qquad \mathbf{(6)}\ \boldsymbol{T \leftarrow B \rightarrow X \leftarrow A \rightarrow Y}\text{: } \boldsymbol{\beta_T\beta_1\alpha_1\alpha_2}.$$

Under linearity and variable standardization, the graphical approach implies that the structural bias of the ANCOVA estimator is expressed as the sum of the products of path coefficients along these open backdoor paths. Then the resulting structural bias is:

$$\boldsymbol{Bias}(\boldsymbol{\gamma_{ANCOVA,3A}})_{\boldsymbol{G}} = \boldsymbol{\alpha_T\alpha_2} + \boldsymbol{\beta_T\beta_2} + (\boldsymbol{\alpha_T\beta_2} + \boldsymbol{\beta_T\alpha_2})(\boldsymbol{\rho} + \boldsymbol{\alpha_1\beta_1}).$$

The first two terms correspond to the core confounding bias through A and through B, whereas the final term captures cross confounding that arises because A and B are correlated (ρ) and because conditioning on the pretest X opens the collider paths implied by the causal structure of X ($\alpha_1\beta_1$).

4.3.1.2. ANCOVA: OLS using FWL Theorem

By the zero conditional mean assumption,

$$\mathrm{Cov}(\varepsilon_X, A) = \mathrm{Cov}(\varepsilon_X, B) = 0,$$

$$\mathrm{Cov}(\varepsilon_T, X) = \mathrm{Cov}(\varepsilon_T, A) = \mathrm{Cov}(\varepsilon_T, B) = 0,$$

$$\mathrm{Cov}(\varepsilon_Y, T) = \mathrm{Cov}(\varepsilon_Y, A) = \mathrm{Cov}(\varepsilon_Y, B) = 0,$$

and all error terms are mutually uncorrelated.

1) Define residuals $\tilde{T}$, $\tilde{A}$, and $\tilde{B}$. By the regression anatomy (FWL) theorem, the ANCOVA population estimator in Scenario 3, Case A is

$$\gamma_{ANCOVA,3A} = \frac{\mathrm{Cov}(Y, \tilde{T})}{\mathrm{Var}(\tilde{T})},$$

where $\tilde{T}$ denotes the residual of T after partialling out X:

$$\tilde{T} = T - \frac{\mathrm{Cov}(T, X)}{\mathrm{Var}(X)}X.$$

In addition, define $\tilde{A}$ and $\tilde{B}$ as the residuals of A and B after partialling out X:

$$\tilde{A} = A - \frac{\mathrm{Cov}(A, X)}{\mathrm{Var}(X)}X,$$

$$\tilde{B} = B - \frac{\mathrm{Cov}(B, X)}{\mathrm{Var}(X)}X.$$

2) Express $\tilde{T}$ in terms of $\tilde{A}$ and $\tilde{B}$. From

$$T = \alpha_T A + \beta_T B + \delta_T X + \varepsilon_T,$$

$$\mathrm{Cov}(T, X) = \mathrm{Cov}(\alpha_T A + \beta_T B + \delta_T X + \varepsilon_T, X) = \delta_T \mathrm{Var}(X) + \alpha_T \mathrm{Cov}(A, X) + \beta_T \mathrm{Cov}(B, X).$$

Plugging this into the definition of $\tilde{T}$ gives

$$\tilde{T} = \alpha_T \tilde{A} + \beta_T \tilde{B} + \varepsilon_T.$$

3) Decompose the numerator $\mathrm{Cov}(Y, \tilde{T})$. From

$$Y = \alpha_2 A + \beta_2 B + \gamma T + \varepsilon_Y,$$

$$\mathrm{Cov}(Y, \tilde{T}) = \mathrm{Cov}(\alpha_2 A + \beta_2 B + \gamma T + \varepsilon_Y, \tilde{T}) = \gamma \mathrm{Var}(\tilde{T}) + \alpha_2 \mathrm{Cov}(A, \tilde{T}) + \beta_2 \mathrm{Cov}(B, \tilde{T}).$$

Using $\tilde{T} = \alpha_T \tilde{A} + \beta_T \tilde{B} + \varepsilon_T$,

$$\mathrm{Cov}(A, \tilde{T}) = \alpha_T \mathrm{Cov}(A, \tilde{A}) + \beta_T \mathrm{Cov}(A, \tilde{B}) = \alpha_T \mathrm{Var}(\tilde{A}) + \beta_T \mathrm{Cov}(A, \tilde{B}),$$

$$\mathrm{Cov}(B, \tilde{T}) = \beta_T \mathrm{Cov}(B, \tilde{B}) + \alpha_T \mathrm{Cov}(B, \tilde{A}) = \beta_T \mathrm{Var}(\tilde{B}) + \alpha_T \mathrm{Cov}(B, \tilde{A}).$$

Therefore,

$$\mathrm{Cov}(Y, \tilde{T}) = \gamma \mathrm{Var}(\tilde{T}) + \alpha_2 \alpha_T \mathrm{Var}(\tilde{A}) + \beta_2 \beta_T \mathrm{Var}(\tilde{B}) + \alpha_T \beta_2 \mathrm{Cov}(B, \tilde{A}) + \beta_T \alpha_2 \mathrm{Cov}(A, \tilde{B}).$$

4) Plug into the regression anatomy formula. These yields

$$\gamma_{\mathrm{ANCOVA},3A} = \gamma + \frac{\alpha_2 \alpha_T \mathrm{Var}(\tilde{A}) + \beta_2 \beta_T \mathrm{Var}(\tilde{B}) + \alpha_T \beta_2 \mathrm{Cov}(B, \tilde{A}) + \beta_T \alpha_2 \mathrm{Cov}(A, \tilde{B})}{\mathrm{Var}(\tilde{T})},$$

and equivalently,

$$\boldsymbol{\gamma_{ANCOVA,3A} = \gamma + \frac{\alpha_2 \alpha_T \mathrm{Var}(\tilde{A}) + \beta_2 \beta_T \mathrm{Var}(\tilde{B}) + (\alpha_T \beta_2 + \beta_T \alpha_2) \mathrm{Cov}(\tilde{A}, \tilde{B})}{\mathrm{Var}(\tilde{T})}}.$$

Note that

$$\mathrm{Cov}(A, \tilde{B}) = \mathrm{Cov}(\tilde{A}, B) = \mathrm{Cov}(\tilde{A}, \tilde{B})$$

when A and B are residualized using the same X. In particular, letting

$$\tilde{A} = A - b_A X, \qquad b_A = \frac{\mathrm{Cov}(A, X)}{\mathrm{Var}(X)},$$

$$\tilde{B} = B - b_B X, \qquad b_B = \frac{\mathrm{Cov}(B, X)}{\mathrm{Var}(X)},$$

so that $\mathrm{Cov}\left(X, \tilde{A}\right) = \mathrm{Cov}\left(X, \tilde{B}\right) = 0$, I have

$$\mathrm{Cov}\left(\tilde{A}, \tilde{B}\right) = \mathrm{Cov}(A, B) - \frac{\mathrm{Cov}(A, X)\mathrm{Cov}(B, X)}{\mathrm{Var}(X)}.$$

5) Residual variances and covariances under standardization. Assuming $\mathrm{Var}(A) = \mathrm{Var}(B) = \mathrm{Var}(X) = 1$ and $\mathrm{Cov}(A, B) = \rho$, first,

$$\mathrm{Var}\left(\tilde{A}\right) = \mathrm{Var}(A) - \frac{\mathrm{Cov}(A, X)^2}{\mathrm{Var}(X)}.$$

Since

$$\mathrm{Cov}(A, X) = \mathrm{Cov}(A, \alpha_1 A + \beta_1 B + \varepsilon_X) = \alpha_1 \mathrm{Var}(A) + \beta_1 \mathrm{Cov}(A, B) = \alpha_1 + \beta_1 \rho,$$

it follows that

$$\mathrm{Var}\left(\tilde{A}\right) = 1 - (\alpha_1 + \beta_1 \rho)^2.$$

Similarly,

$$\mathrm{Var}\left(\tilde{B}\right) = \mathrm{Var}(B) - \frac{\mathrm{Cov}(B, X)^2}{\mathrm{Var}(X)},$$

and since

$$\mathrm{Cov}(B, X) = \mathrm{Cov}(B, \alpha_1 A + \beta_1 B + \varepsilon_X) = \alpha_1 \mathrm{Cov}(B, A) + \beta_1 \mathrm{Var}(B) = \beta_1 + \alpha_1 \rho,$$

I obtain

$$\mathrm{Var}\left(\tilde{B}\right) = 1 - (\beta_1 + \alpha_1 \rho)^2.$$

Thus,

$$\mathrm{Cov}\left(\tilde{A}, \tilde{B}\right) = \rho - (\alpha_1 + \beta_1 \rho)(\beta_1 + \alpha_1 \rho).$$

6) Finally, $\mathrm{Var}\left(\tilde{T}\right)$ can be written as

$$\mathrm{Var}\left(\tilde{T}\right) = \mathrm{Var}(T) - \frac{\mathrm{Cov}(T, X)^2}{\mathrm{Var}(X)}.$$

Under standardization, using $\mathrm{Var}(T) = \mathrm{Var}(X) = 1$ and

$$\mathrm{Cov}(T, X) = \delta_T + \alpha_T(\alpha_1 + \beta_1\rho) + \beta_T(\beta_1 + \alpha_1\rho),$$

I obtain

$$\mathrm{Var}(\tilde{T}) = 1 - \{\delta_T + \alpha_T(\alpha_1 + \beta_1\rho) + \beta_T(\beta_1 + \alpha_1\rho)\}^2.$$

Thus, under standardization,

$$\boldsymbol{\gamma_{ANCOVA,3A}}$$

$$= \gamma$$

$$+ \frac{\boldsymbol{\alpha_2\alpha_T\{1 - (\alpha_1 + \beta_1\rho)^2\} + \beta_2\beta_T\{1 - (\beta_1 + \alpha_1\rho)^2\} + (\alpha_T\beta_2 + \beta_T\alpha_2)\{\rho - (\alpha_1 + \beta_1\rho)(\beta_1 + \alpha_1\rho)\}}}{\boldsymbol{1 - \{\delta_T + \alpha_T(\alpha_1 + \beta_1\rho) + \beta_T(\beta_1 + \alpha_1\rho)\}^2}}$$

where the terms added to γ constitute $\boldsymbol{Bias(\gamma_{ANCOVA,3A})_{OLS}}$.

7) Unbiasedness conditions

Assuming $\mathrm{Var}(\tilde{T}) > 0$, the ANCOVA estimator under Scenario 3, Case A equals the causal treatment effect, $\gamma_{\mathrm{ANCOVA},3A} = \gamma$, whenever at least one of the following conditions holds:

i) $\alpha_T = \beta_T = 0$ (no $A \to T$ and no $B \to T$).

ii) $\alpha_2 = \beta_2 = 0$ (no $A \to Y$ and no $B \to Y$).

iii) $\mathrm{Var}(\tilde{A}) = 0$ and $\mathrm{Var}(\tilde{B}) = 0$, that is, after partialling out X, neither A nor B has any remaining residual variation. In this setting, X functions as a perfect proxy for both confounders simultaneously, and it follows immediately that $\mathrm{Cov}(\tilde{A}, \tilde{B}) = 0$ because $\tilde{A}$ and $\tilde{B}$ collapse to constants.

Therefore, $Bias(\gamma_{ANCOVA,3A})_G$ captures the core confounding component identified by the graphical approach, whereas $Bias(\gamma_{ANCOVA,3A})_{OLS}$ expresses the total ANCOVA structural bias with the residualization-induced scaling implied by the FWL theorem. Under Scenario 3, Case A, the data-generating DAG extends the one confounder setting by adding a second unobserved

confounder B and allowing A and B to be correlated (denote $\rho = \mathrm{Cov}(A, B)$). Under ANCOVA, the analytic adjustment set remains $\{X\}$, so the graphical analysis focuses on the open backdoor structure induced by the two confounders and their correlation. Under variable standardization, applying the path-tracing rule to the open backdoor paths yields the graphical structural bias expression:

$$\boldsymbol{Bias}(\boldsymbol{\gamma}_{\boldsymbol{ANCOVA,3A}})_{\boldsymbol{G}} = \boldsymbol{\alpha}_{\boldsymbol{T}}\boldsymbol{\alpha}_{\boldsymbol{2}} + \boldsymbol{\beta}_{\boldsymbol{T}}\boldsymbol{\beta}_{\boldsymbol{2}} + (\boldsymbol{\alpha}_{\boldsymbol{T}}\boldsymbol{\beta}_{\boldsymbol{2}} + \boldsymbol{\beta}_{\boldsymbol{T}}\boldsymbol{\alpha}_{\boldsymbol{2}})(\boldsymbol{\rho} + \boldsymbol{\alpha}_{\boldsymbol{1}}\boldsymbol{\beta}_{\boldsymbol{1}}),$$

where the first two terms represent the direct confounding components through $A$and B, and the final term captures cross confounding that arises from the correlation between $A$and Btogether with the A and B mixture structure of the pretest.

The regression anatomy result clarifies why the two-confounder case differs from Scenario 2. With two unobserved confounders, the OLS-based ANCOVA bias is not obtained by scaling a single core confounding product by one residual variance ratio. Instead, the OLS expression scales confounding contributions through the residual covariance structure among A and B after partialling out X through $\mathrm{Var}(\tilde{A})$, $\mathrm{Var}(\tilde{B})$, and $\mathrm{Cov}(\tilde{A}, \tilde{B})$, rather than through a single residual variance ratio. Even so, the graphical decomposition delivers two practically interpretable sufficient conditions for zero bias, because these correspond to eliminating the remaining open backdoor structure in the DAG: $\alpha_T = \beta_T = 0$(neither unobserved confounder affects treatment assignment) or $\alpha_2 = \beta_2 = 0$ (neither unobserved confounder affects the outcome). By contrast, perfect proxy type restrictions are not recoverable from the DAG alone, because they are strength or measurement conditions rather than purely structural features of the graph. Under Scenario 3, these proxy strength features enter only through the regression anatomy by determining how the core confounding components are scaled via $\mathrm{Var}(\tilde{A})$, $\mathrm{Var}(\tilde{B})$, and $\mathrm{Cov}(\tilde{A}, \tilde{B})$.

4.3.1.3. Change Score Analysis: Graphical Approach

Scenario 3, Case A extends the one confounder setting by adding a second unobserved confounder B (in addition to A) and allowing A and B to be correlated (denote $\rho = \rho_A \rho_B$). Because the change score analytic model is still a simple regression of C on T, neither X nor the unobserved confounders are conditioned on, so the additional confounder B and the $A - B$ association generate additional open backdoor paths from T to C. Listing only the open backdoor paths with their path-tracing products under standardization, I have:

Confounding via the pretest X

1) $T \leftarrow X \rightarrow C$: $-\delta_T$

Confounding via A

2) $T \leftarrow A \rightarrow Y \rightarrow C$: $\alpha_T \alpha_2$

3) $T \leftarrow A \rightarrow X \rightarrow C$: $\alpha_T \alpha_1$

Confounding via B

4) $T \leftarrow B \rightarrow Y \rightarrow C$: $\beta_T \beta_2$

5) $T \leftarrow B \rightarrow X \rightarrow C$: $-\beta_T \beta_1$

Confounding via the A–B correlation (ρ)

6) $T \leftarrow A \leftarrow U \rightarrow B \rightarrow Y \rightarrow C$: $\alpha_T \rho \beta_2$

7) $T \leftarrow B \leftarrow U \rightarrow A \rightarrow Y \rightarrow C$: $\beta_T \rho \alpha_2$

8) $T \leftarrow A \leftarrow U \rightarrow B \rightarrow X \rightarrow C$: $-\alpha_T \rho \beta_1$

9) $T \leftarrow B \leftarrow U \rightarrow A \rightarrow X \rightarrow C$: $-\beta_T \rho \alpha_1$

Confounding via X together with the two confounders

10) $T \leftarrow X \leftarrow A \rightarrow Y \rightarrow C$: $\delta_T \alpha_1 \alpha_2$

11) $T \leftarrow X \leftarrow B \rightarrow Y \rightarrow C$: $\delta_T \beta_1 \beta_2$

12) $T \leftarrow X \leftarrow A \leftarrow U \rightarrow B \rightarrow Y \rightarrow C: \delta_T \alpha_1 \rho \beta_2$

13) $T \leftarrow X \leftarrow B \leftarrow U \rightarrow A \rightarrow Y \rightarrow C: \delta_T \beta_1 \rho \alpha_2$

Summing up these path-tracing products yields the structural bias of the change score estimator under Scenario 3, Case A:

$$\boldsymbol{Bias}(\boldsymbol{\gamma_{Change,3A}})_{\boldsymbol{G}}$$

$$= \boldsymbol{\delta_T}[\boldsymbol{\alpha_1\alpha_2} + \boldsymbol{\beta_1\beta_2} - \mathbf{1} + \boldsymbol{\rho}(\boldsymbol{\alpha_1\beta_2} + \boldsymbol{\alpha_2\beta_1})] + \boldsymbol{\alpha_T}[(\boldsymbol{\alpha_2} - \boldsymbol{\alpha_1}) + \boldsymbol{\rho}(\boldsymbol{\beta_2} - \boldsymbol{\beta_1})]$$

$$+ \boldsymbol{\beta_T}[(\boldsymbol{\beta_2} - \boldsymbol{\beta_1}) + \boldsymbol{\rho}(\boldsymbol{\alpha_2} - \boldsymbol{\alpha_1})].$$

4.3.1.4. Change Score Analysis: OLS

Under Scenario 3, Case A, the change score estimator is expressed as

$$\boldsymbol{\gamma_{Change,3A}} = \mathrm{Cov}(Y,T) - \mathrm{Cov}(X,T)$$

$$= \gamma + \boldsymbol{\delta_T}[\boldsymbol{\alpha_1\alpha_2} + \boldsymbol{\beta_1\beta_2} - \mathbf{1} + \boldsymbol{\rho}(\boldsymbol{\alpha_1\beta_2} + \boldsymbol{\alpha_2\beta_1})] + \boldsymbol{\alpha_T}[(\boldsymbol{\alpha_2} - \boldsymbol{\alpha_1}) + \boldsymbol{\rho}(\boldsymbol{\beta_2} - \boldsymbol{\beta_1})] +$$

$\boldsymbol{\beta_T}[(\boldsymbol{\beta_2} - \boldsymbol{\beta_1}) + \boldsymbol{\rho}(\boldsymbol{\alpha_2} - \boldsymbol{\alpha_1})]$ where the terms added to γ constitute $\boldsymbol{Bias}(\boldsymbol{\gamma_{Change,3A}})_{\boldsymbol{OLS}}$.

Given the required covariance terms are derived as follows (under standardization $\mathrm{Var}(A) = \mathrm{Var}(B) = \mathrm{Var}(X) = \mathrm{Var}(T) = 1$, $\mathrm{Cov}(A,B) = \rho$, and the zero conditional mean assumptions)

1) Derive $\mathrm{Cov}(X,T)$

Using $T = \delta_T X + \alpha_T A + \beta_T B + \varepsilon_T$,

$$\mathrm{Cov}(X,T) = \mathrm{Cov}(X, \alpha_T A + \beta_T B + \delta_T X + \varepsilon_T),$$

$$\mathrm{Cov}(X,T) = \delta_T \mathrm{Var}(X) + \alpha_T \mathrm{Cov}(X,A) + \beta_T \mathrm{Cov}(X,B).$$

Since $X = \alpha_1 A + \beta_1 B + \varepsilon_X$,

$$\mathrm{Cov}(X,A) = \alpha_1 \mathrm{Var}(A) + \beta_1 \mathrm{Cov}(B,A) = \alpha_1 + \beta_1 \rho,$$

$$\mathrm{Cov}(X,B) = \beta_1 \mathrm{Var}(B) + \alpha_1 \mathrm{Cov}(A,B) = \beta_1 + \alpha_1 \rho.$$

Therefore,

$$\mathrm{Cov}(X,T) = \delta_T + \alpha_T(\alpha_1 + \beta_1 \rho) + \beta_T(\beta_1 + \alpha_1 \rho).$$

2) Derive $\mathrm{Cov}(Y,T)$

Using $Y = \alpha_2 A + \beta_2 B + \gamma T + \varepsilon_Y$,

$$\mathrm{Cov}(Y,T) = \mathrm{Cov}(\alpha_2 A + \beta_2 B + \gamma T + \varepsilon_Y, T),$$

$$\mathrm{Cov}(Y,T) = \gamma \mathrm{Var}(T) + \alpha_2 \mathrm{Cov}(A,T) + \beta_2 \mathrm{Cov}(B,T),$$

$$\mathrm{Cov}(Y,T) = \gamma + \alpha_2 \mathrm{Cov}(A,T) + \beta_2 \mathrm{Cov}(B,T).$$

Compute $\mathrm{Cov}(A,T)$ and $\mathrm{Cov}(B,T)$ from $T = \alpha_T A + \beta_T B + \delta_T X + \varepsilon_T$:

$$\mathrm{Cov}(A,T) = \delta_T \mathrm{Cov}(A,X) + \alpha_T \mathrm{Var}(A) + \beta_T \mathrm{Cov}(A,B) = \delta_T \mathrm{Cov}(A,X) + \alpha_T + \beta_T \rho,$$

$$\mathrm{Cov}(B,T) = \delta_T \mathrm{Cov}(B,X) + \beta_T \mathrm{Var}(B) + \alpha_T \mathrm{Cov}(A,B) = \delta_T \mathrm{Cov}(B,X) + \beta_T + \alpha_T \rho.$$

Since $X = \alpha_1 A + \beta_1 B + \varepsilon_X$,

$$\mathrm{Cov}(A,X) = \alpha_1 + \beta_1 \rho, \qquad \mathrm{Cov}(B,X) = \beta_1 + \alpha_1 \rho,$$

so

$$\mathrm{Cov}(A,T) = \delta_T(\alpha_1 + \beta_1 \rho) + \alpha_T + \beta_T \rho, \qquad \mathrm{Cov}(B,T) = \delta_T(\beta_1 + \alpha_1 \rho) + \beta_T + \alpha_T \rho.$$

Substituting,

$$\mathrm{Cov}(Y,T) = \gamma + \alpha_2[\delta_T(\alpha_1 + \beta_1 \rho) + \alpha_T + \beta_T \rho] + \beta_2[\delta_T(\beta_1 + \alpha_1 \rho) + \beta_T + \alpha_T \rho].$$

Therefore, $Bias(\gamma_{Change,3A})_G$ remains identical to $Bias(\gamma_{Change,3A})_{OLS}$. Relative to Scenario 2, adding a second unobserved confounder B and allowing A and B to be correlated introduces additional confounding components through B itself and through the cross confounding induced by the A–B association. As a result, the familiar Scenario 2 "common trend" condition $\alpha_2 = \alpha_1$ generalizes to a pair of coupled time invariant confounding conditions that explicitly incorporate ρ:

$$\alpha_2 + \rho\beta_2 = \alpha_1 + \rho\beta_1 \text{ and } \beta_2 + \rho\alpha_2 = \beta_1 + \rho\alpha_1,$$

together with $\delta_T = 0$ (i.e., no dynamic causal relationship between the pretest and treatment). Substantively, these generalized conditions state that, once the correlation between the two

unobserved confounders is taken into account, each confounder's total effect on the untreated posttest matches its effect on the pretest. A simpler but stronger sufficient version sets $\alpha_2 = \alpha_1$ and $\beta_2 = \beta_1$, which automatically satisfies the coupled conditions. Alternatively, unbiasedness also holds under no confounding of treatment assignment by either unobserved confounder, $\alpha_T = \beta_T = 0$. In addition, unbiasedness is also technically possible through exact bias offsetting, which occurs when $\delta_T [\alpha_1\alpha_2 + \beta_1\beta_2 - 1 + \rho(\alpha_1\beta_2 + \alpha_2\beta_1)] + \alpha_T [(\alpha_2 - \alpha_1) + \rho(\beta_2 - \beta_1)] + \beta_T [(\beta_2 - \beta_1) + \rho(\alpha_2 - \alpha_1)] = 0$.

4.3.2. Case B: When the pretest affects the posttest

4.3.2.1. ANCOVA: Graphical Approach

Under Scenario 3, Case B, the data generating DAG extends Case A by adding a direct pretest to posttest effect $X \rightarrow Y$ with coefficient τ. As in Scenarios 1 and 2, introducing the additional edge $X \rightarrow Y$ creates additional potential backdoor paths from T to Y that run through X. However, because ANCOVA conditions on X, these X mediated backdoor paths are blocked and therefore do not contribute to structural bias in the ANCOVA coefficient on T. For example, conditioning on X blocks paths such as $T \leftarrow X \rightarrow Y$, $T \leftarrow A \rightarrow X \rightarrow Y$, *and* $T \leftarrow B \rightarrow X \rightarrow Y$. Thus, the direct effect X → Y does not introduce additional open backdoor paths for ANCOVA in this case. Consequently, the set of *open* backdoor paths relevant for bias under ANCOVA is the same as in Scenario 3, Case A: Under variable standardization, applying the path-tracing rule to these open backdoor paths yields the same graphical structural bias expression as in Case A:

$$\boldsymbol{Bias}(\boldsymbol{\gamma_{ANCOVA,3B}})_{\boldsymbol{G}} = \boldsymbol{\alpha_T\alpha_2 + \beta_T\beta_2 + (\alpha_T\beta_2 + \beta_T\alpha_2)(\rho + \alpha_1\beta_1)}.$$

Thus, although Case B adds a direct carryover effect from the pretest to the posttest, that additional causal link does not alter the open backdoor structure that remains after conditioning on

X, and it therefore does not change the graphical characterization of ANCOVA structural bias relative to Scenario 3, Case A.

4.3.2.2. ANCOVA: OLS using FWL Theorem

By the zero conditional mean assumption,

$$Cov(\varepsilon_X, A) = Cov(\varepsilon_X, B) = 0,$$

$$Cov(\varepsilon_T, X) = Cov(\varepsilon_T, A) = Cov(\varepsilon_T, B) = 0,$$

$$Cov(\varepsilon_Y, T) = Cov(\varepsilon_Y, A) = Cov(\varepsilon_Y, B) = 0,$$

and all error terms are mutually uncorrelated.

1) Define residuals $\tilde{T}, \tilde{A}, and\ \tilde{B}$. By the regression anatomy (FWL) theorem, the ANCOVA population estimator in Scenario 3, Case B is

$$\gamma_{ANCOVA,3B} = \frac{Cov(Y, \tilde{T})}{Var(\tilde{T})},$$

where $\tilde{T}$ denotes the residual of T after partialling out X:

$$\tilde{T} = T - \frac{Cov(T, X)}{Var(X)} X.$$

In addition, define $\tilde{A}$ and $\tilde{B}$ as the residuals of A and B after partialling out X:

$$\tilde{A} = A - \frac{Cov(A, X)}{Var(X)} X,$$

$$\tilde{B} = B - \frac{Cov(B, X)}{Var(X)} X.$$

2) Express $\tilde{T}$ in terms of $\tilde{A}$ and $\tilde{B}$. From

$$T = \alpha_T\, A + \beta_T\, B + \delta_T\, X + \varepsilon_T,$$

$$Cov(T, X) = \delta_T\, Var(X) + \alpha_T\, Cov(A, X) + \beta_T\, Cov(B, X).$$

Plugging this into the definition of $\tilde{T}$ gives

$$\tilde{T} = \alpha_T \tilde{A} + \beta_T \tilde{B} + \varepsilon_T.$$

3) Decompose the numerator $Cov(Y, \tilde{T})$. From

$$Y = \alpha_2 A + \beta_2 B + \tau X + \gamma T + \varepsilon_Y,$$

$$Cov(Y, \tilde{T}) = \tau\, Cov(X, \tilde{T}) + \gamma\, Var(\tilde{T}) + \alpha_2\, Cov(A, \tilde{T}) + \beta_2\, Cov(B, \tilde{T}).$$

Note that $\mathrm{Cov}(\mathrm{X}, \widetilde{\mathrm{T}}) = 0$ by construction because $\widetilde{\mathrm{T}}$ is the population OLS residual from projecting T on X. Thus, the τX term does not affect the partial regression coefficient on T.

Using $\tilde{T} = \alpha_T \tilde{A} + \beta_T \tilde{B} + \varepsilon_T$, we have

$$Cov(A, \tilde{T}) = \alpha_T\, Var(\tilde{A}) + \beta_T\, Cov(\tilde{A}, \tilde{B}),$$

$$Cov(B, \tilde{T}) = \beta_T\, Var(\tilde{B}) + \alpha_T\, Cov(\tilde{A}, \tilde{B}).$$

Therefore,

$$Cov(Y, \tilde{T}) = \gamma\, Var(\tilde{T}) + \alpha_2\alpha_T\, Var(\tilde{A}) + \beta_2\beta_T\, Var(\tilde{B}) + (\alpha_T\beta_2 + \beta_T\alpha_2) Cov(\tilde{A}, \tilde{B}).$$

4) Plug into the regression anatomy formula. Then,

$$\boldsymbol{\gamma_{ANCOVA,3B}} = \gamma + \frac{\boldsymbol{\alpha_2\alpha_T\, Var(\widetilde{A}) + \beta_2\beta_T\, Var(\widetilde{B}) + (\alpha_T\beta_2 + \beta_T\alpha_2) Cov(\widetilde{A}, \widetilde{B})}}{\boldsymbol{Var(\widetilde{T})}}.$$

Note that $Cov(A, \tilde{B}) = Cov(\tilde{A}, B) = Cov(\tilde{A}, \tilde{B})$ when A and B are residualized using the same *X*.

5) Since residual variances and covariance are the same as Case A,

$$\boldsymbol{\gamma_{ANCOVA,3B}}$$

$$= \gamma$$

$$+ \frac{\boldsymbol{\alpha_2\alpha_T\{1 - (\alpha_1 + \beta_1\rho)^2\} + \beta_2\beta_T\{1 - (\beta_1 + \alpha_1\rho)^2\} + (\alpha_T\beta_2 + \beta_T\alpha_2)\{\rho - (\alpha_1 + \beta_1\rho)(\beta_1 + \alpha_1\rho)\}}}{\boldsymbol{1 - \{\delta_T + \alpha_T(\alpha_1 + \beta_1\rho) + \beta_T(\beta_1 + \alpha_1\rho)\}^2}}$$

where the terms added to γ consititute $\boldsymbol{Bias(\gamma_{ANCOVA,3B})_{OLS}}$, which is the same as $\boldsymbol{Bias(\gamma_{ANCOVA,3A})_{OLS}}$.

Therefore, under Scenario 3, Case B, adding a direct pretest to posttest effect $X \rightarrow Y$ with coefficient τ does not change the ANCOVA structural bias or the unbiasedness conditions relative

to Case A. From the graphical approach, the additional edge $X \rightarrow Y$can generate new potential noncausal routes from T to Y that run through X, but because ANCOVA conditions on X, those X mediated paths do not remain open and therefore do not contribute to bias. The remaining open structure is the same as in Scenario 3, Case A, thus the same practically interpretable sufficient zero bias conditions apply as in Case A: bias is zero if neither confounder affects treatment assignment ($\alpha_T = \beta_T = 0$)or if neither confounder affects the posttest outcome ($\alpha_2 = \beta_2 = 0$). The regression anatomy makes the invariance to τ explicit, because $\gamma_{\text{ANCOVA}} = \text{Cov}(Y, \tilde{T})/\text{Var}(\tilde{T})$ with $\tilde{T}$, the residual from regressing Ton X, and $\text{Cov}(X, \tilde{T}) = 0$ implies the τX component of Ycannot enter $\text{Cov}(Y, \tilde{T})$. What remains is the core confounding component identified by the graph, scaled by the residual covariance structure of the omitted confounders after partialling out X, through $\text{Var}(\tilde{A})$, $\text{Var}(\tilde{B})$, and $\text{Cov}(\tilde{A}, \tilde{B})$, rather than through a single residual variance ratio.

4.3.2.3. Change Score Analysis: Graphical Approach

Under Scenario 3, Case B extends Case A by adding a direct pretest to posttest effect $X \rightarrow Y$ with coefficient τ, in a setting with two unobserved confounders $A$and B that may be correlated and jointly induce confounding between T and the change score C. Relative to Case A, introducing $X \rightarrow Y$adds five additional τindexed open backdoor paths from $T$to C, increasing the total number of open backdoor paths from 13 to 18, because existing backdoor connections that reach X can now continue through the new link $X \rightarrow Y$ before entering C via $Y \rightarrow C$.

Confounding via the pretest *X*

1) $T \leftarrow X \rightarrow C$: $-\delta_T$

2) $T \leftarrow X \rightarrow Y \rightarrow C$: $\delta_T \tau$

Confounding via *A*

3) $T \leftarrow A \rightarrow Y \rightarrow C$: $\alpha_T \alpha_2$

4) $T \leftarrow A \rightarrow X \rightarrow C$: $-\alpha_T \alpha_1$

5) $T \leftarrow A \rightarrow X \rightarrow Y \rightarrow C$: $\alpha_T \alpha_1 \tau$

Confounding via *B*

6) $T \leftarrow B \rightarrow Y \rightarrow C$: $\beta_T \beta_2$

7) $T \leftarrow B \rightarrow X \rightarrow C$: $-\beta_T \beta_1$

8) $T \leftarrow B \rightarrow X \rightarrow Y \rightarrow C$: $\beta_T \beta_1 \tau$

Confounding via the *A*–*B* correlation ($A \leftarrow U \rightarrow B$ or $B \leftarrow U \rightarrow A$)

9) $T \leftarrow A \leftarrow U \rightarrow B \rightarrow Y \rightarrow C$: $\alpha_T \rho \beta_2$

10) $T \leftarrow A \leftarrow U \rightarrow B \rightarrow X \rightarrow C$: $-\alpha_T \rho \beta_1$

11) $T \leftarrow A \leftarrow U \rightarrow B \rightarrow X \rightarrow Y \rightarrow C$: $\alpha_T \rho \beta_1 \tau$

12) $T \leftarrow B \leftarrow U \rightarrow A \rightarrow Y \rightarrow C$: $\beta_T \rho \alpha_2$

13) $T \leftarrow B \leftarrow U \rightarrow A \rightarrow X \rightarrow C$: $-\beta_T \rho \alpha_1$

14) $T \leftarrow B \leftarrow U \rightarrow A \rightarrow X \rightarrow Y \rightarrow C$: $\beta_T \rho \alpha_1 \tau$

Confounding via *X* and the two confounders

15) $T \leftarrow X \leftarrow A \rightarrow Y \rightarrow C$: $\delta_T \alpha_1 \alpha_2$

16) $T \leftarrow X \leftarrow B \rightarrow Y \rightarrow C$: $\delta_T \beta_1 \beta_2$

17) $T \leftarrow X \leftarrow A \leftarrow U \rightarrow B \rightarrow Y \rightarrow C$: $\delta_T \alpha_1 \rho \beta_2$

18) $T \leftarrow X \leftarrow B \leftarrow U \rightarrow A \rightarrow Y \rightarrow C$: $\delta_T \beta_1 \rho \alpha_2$

Summing these components yields:

$$\boldsymbol{Bias}(\boldsymbol{\gamma_{Change,3B}})_{\boldsymbol{G}}$$

$$= \boldsymbol{\delta_T}[\boldsymbol{\alpha_1\alpha_2} + \boldsymbol{\beta_1\beta_2} - \mathbf{1} + \boldsymbol{\rho}(\boldsymbol{\alpha_1\beta_2} + \boldsymbol{\alpha_2\beta_1}) + \boldsymbol{\tau}]$$

$$+ \boldsymbol{\alpha_T}[(\boldsymbol{\alpha_2} - \boldsymbol{\alpha_1}) + \boldsymbol{\rho}(\boldsymbol{\beta_2} - \boldsymbol{\beta_1}) + \boldsymbol{\tau}(\boldsymbol{\alpha_1} + \boldsymbol{\rho\beta_1})]$$

$$+ \boldsymbol{\beta_T}[(\boldsymbol{\beta_2} - \boldsymbol{\beta_1}) + \boldsymbol{\rho}(\boldsymbol{\alpha_2} - \boldsymbol{\alpha_1}) + \boldsymbol{\tau}(\boldsymbol{\beta_1} + \boldsymbol{\rho\alpha_1})].$$

4.3.2.4. Change Score Analysis: OLS

Under Scenario 3, Case B, with $Y = \alpha_2 A + \beta_2 B + \tau X + \gamma T + \varepsilon_Y$,

$$\mathrm{Cov}(Y,T) = \tau\,\mathrm{Cov}(X,T) + \gamma + \alpha_2\,\mathrm{Cov}(A,T) + \beta_2\,\mathrm{Cov}(B,T),$$

so

$$\gamma_{Change,3B} = \mathrm{Cov}(Y,T) - \mathrm{Cov}(X,T)$$

$$= \gamma + (\tau - 1)\mathrm{Cov}(X,T) + \alpha_2\mathrm{Cov}(A,T) + \beta_2\mathrm{Cov}(B,T).$$

Using $\mathrm{Cov}(X,T) = \alpha_1\mathrm{Cov}(A,T) + \beta_1\mathrm{Cov}(B,T)$,

$\gamma_{\mathrm{Change},3B} = \gamma + [\alpha_2 + (\tau - 1)\alpha_1]\mathrm{Cov}(A,T) + [\beta_2 + (\tau - 1)\beta_1]\mathrm{Cov}(B,T).$

The treatment equation is unchanged, so the covariance expressions for $\mathrm{Cov}(A,T)$ and $\mathrm{Cov}(B,T)$ are the same as in Case A. Thus,

$$\boldsymbol{\gamma_{Change,3B}} = \gamma + \boldsymbol{\delta_T}[\boldsymbol{\alpha_1\alpha_2} + \boldsymbol{\beta_1\beta_2} - \mathbf{1} + \boldsymbol{\rho}(\boldsymbol{\alpha_1\beta_2} + \boldsymbol{\alpha_2\beta_1}) + \boldsymbol{\tau}]$$

$$+ \boldsymbol{\alpha_T}[(\boldsymbol{\alpha_2} - \boldsymbol{\alpha_1}) + \boldsymbol{\rho}(\boldsymbol{\beta_2} - \boldsymbol{\beta_1}) + \boldsymbol{\tau}(\boldsymbol{\alpha_1} + \boldsymbol{\rho\beta_1})]$$

$$+ \boldsymbol{\beta_T}[(\boldsymbol{\beta_2} - \boldsymbol{\beta_1}) + \boldsymbol{\rho}(\boldsymbol{\alpha_2} - \boldsymbol{\alpha_1}) + \boldsymbol{\tau}(\boldsymbol{\beta_1} + \boldsymbol{\rho\alpha_1})]$$

where the terms added to γ constitute $\boldsymbol{Bias}(\boldsymbol{\gamma_{Change,3B}})_{\boldsymbol{OLS}}$.

Therefore, $Bias(\gamma_{Change,3B})_G$ remains identical to $Bias(\gamma_{Change,3B})_{OLS}$. A practically interpretable unbiasedness condition begins with $\delta_T = 0$, meaning there is no dynamic causal relationship from the pretest to treatment assignment. Given $\delta_T = 0$, unbiasedness additionally requires the generalized common trend (time invariant confounding) condition:

$$\alpha_2 + \tau\alpha_1 + \rho(\beta_2 + \tau\beta_1) = \alpha_1 + \rho\beta_1 \text{ and } \beta_2 + \tau\beta_1 + \rho(\alpha_2 + \tau\alpha_1) = \beta_1 + \rho\alpha_1,$$

which incorporate both the A–B correlation ρ and the pretest to posttest carryover τ when equating each confounder's total effect on the untreated posttest with its effect on the pretest. A simplified but stronger sufficient version sets $\alpha_2 = \alpha_1$, $\beta_2 = \beta_1$, and $\tau = 0$. Alternatively, unbiasedness also holds if $\alpha_T = 0$ and $\beta_T = 0$, so neither unobserved confounder affects treatment assignment. In addition, unbiasedness is also technically possible through exact bias offsetting, which occurs when $\delta_T [\alpha_1\alpha_2 + \beta_1\beta_2 - 1 + \rho(\alpha_1\beta_2 + \alpha_2\beta_1) + \tau] + \alpha_T [(\alpha_2 - \alpha_1) + \rho(\beta_2 - \beta_1) + \tau(\alpha_1 + \rho\beta_1)] + \beta_T [(\beta_2 - \beta_1) + \rho(\alpha_2 - \alpha_1) + \tau(\beta_1 + \rho\alpha_1)] = 0$.

4.3.3. Case C: When the pretest affects the posttest through a mediator

4.3.3.1. ANCOVA: Graphical Approach

Under Scenario 3, Case C, the data generating DAG extends Case B by adding an unobserved mediator M between waves (with $A \rightarrow M$ and $M \rightarrow Y$, and with treatment affecting the mediator through $T \rightarrow M$). The ANCOVA analytic model continues to condition only on the pretest X, so the implied adjustment set remains $\{X\}$. As in Scenario 3, Case A and Case B, I focus only on the open backdoor paths from T to Y and their path-tracing products.

Relative to Case B, all previously open backdoor paths remain open, and the presence of M introduces additional open backdoor paths that run through M. In particular, Case C adds the following open backdoor paths (and associated path-tracing products):

7) $\boldsymbol{T \leftarrow A \rightarrow M \rightarrow Y}$: $\boldsymbol{\alpha_T\alpha_M\eta}$,

8) $\boldsymbol{T \leftarrow B \leftarrow U \rightarrow A \rightarrow M \rightarrow Y}$: $\boldsymbol{\beta_T\rho\alpha_M\eta}$,

9) $\boldsymbol{T \leftarrow B \rightarrow X \leftarrow A \rightarrow M \rightarrow Y}$: $\boldsymbol{\beta_T\beta_1\alpha_1\alpha_M\eta}$.

In addition, because the treatment effect is defined as the direct effect of T on Y while M is omitted from the ANCOVA analytic model, the causal mediation path contributes structural bias as omitted mediator bias:

$$\boldsymbol{T \rightarrow M \rightarrow Y : \zeta\eta.}$$

Summing the open backdoor contributions together with this omitted mediator term yields the structural bias identified by the graphical approach. Under standardization, this gives:

$$\boldsymbol{Bias(\gamma_{ANCOVA,3C})_G = \zeta\eta + \alpha_T(\alpha_2 + \alpha_M\eta) + \beta_T\beta_2 + (\alpha_T\beta_2 + \beta_T\alpha_2 + \beta_T\alpha_M\eta)(\rho + \alpha_1\beta_1).}$$

This expression highlights that Case C differs from Case B by adding (i) new confounding terms transmitted through the mediator M and (ii) the additional omitted mediator component $\zeta\eta$, while preserving the same two confounder and cross confounding structure already present in Scenario 3.

4.3.3.2. ANCOVA: OLS using FWL Theorem

By the zero conditional mean assumption,

$\mathrm{Cov}(\varepsilon_X, A) = \mathrm{Cov}(\varepsilon_X, B) = 0,$

$\mathrm{Cov}(\varepsilon_T, X) = \mathrm{Cov}(\varepsilon_T, A) = \mathrm{Cov}(\varepsilon_T, B) = 0,$

$\mathrm{Cov}(\varepsilon_M, X) = \mathrm{Cov}(\varepsilon_M, T) = \mathrm{Cov}(\varepsilon_M, A) = 0,$

$\mathrm{Cov}(\varepsilon_Y, X) = \mathrm{Cov}(\varepsilon_Y, T) = \mathrm{Cov}(\varepsilon_Y, M) = \mathrm{Cov}(\varepsilon_Y, A) = \mathrm{Cov}(\varepsilon_Y, B) = 0,$

and all error terms are mutually uncorrelated.

1) Define residuals $\tilde{T}$, $\tilde{A}$, and $\tilde{B}$

   By the regression anatomy identity, the ANCOVA coefficient on T is

$$\gamma_{ANCOVA,3C} = \frac{\mathrm{Cov}(Y, \tilde{T})}{\mathrm{Var}(\tilde{T})},$$

   where $\tilde{T}$ denotes the residual of T after partialing out X. Define

$$\tilde{T} = T - \frac{\mathrm{Cov}(T, X)}{\mathrm{Var}(X)}X, \qquad \tilde{A} = A - \frac{\mathrm{Cov}(A, X)}{\mathrm{Var}(X)}X, \qquad \tilde{B} = B - \frac{\mathrm{Cov}(B, X)}{\mathrm{Var}(X)}X.$$

2) Express $\tilde{T}$ in terms of $\tilde{A}$ and $\tilde{B}$

   From $T = \alpha_T A + \beta_T B + \delta_T X + \varepsilon_T$,

$$\mathrm{Cov}(T, X) = \delta_T \mathrm{Var}(X) + \alpha_T \mathrm{Cov}(A, X) + \beta_T \mathrm{Cov}(B, X).$$

Plugging this into the residual definition of $\tilde{T}$ gives

$$\tilde{T} = T - \frac{\mathrm{Cov}(T, X)}{\mathrm{Var}(X)} X = \alpha_T \tilde{A} + \beta_T \tilde{B} + \varepsilon_T.$$

3) Reduce Y to the function of X, T, A, and B

Substitute the mediator equation $M = \alpha_M A + \delta_M X + \zeta T + \varepsilon_M$ into the outcome equation to obtain

$$Y = (\tau + \eta\delta_M)X + (\gamma + \eta\zeta)T + (\alpha_2 + \eta\alpha_M)A + \beta_2 B + (\varepsilon_Y + \eta\varepsilon_M).$$

4) Shorthand notation and numerator decomposition

Define

$$\tau^* = \tau + \eta\delta_M, \qquad \gamma^* = \gamma + \eta\zeta, \qquad \alpha_2^* = \alpha_2 + \eta\alpha_M, \qquad \varepsilon_Y^* = \eta\varepsilon_M + \varepsilon_Y,$$

so that

$$Y = \alpha_2^* A + \beta_2 B + \tau^* X + \gamma^* T + \varepsilon_Y^*.$$

Decompose the numerator:

$$\mathrm{Cov}(Y, \tilde{T}) = \tau^* \mathrm{Cov}(X, \tilde{T}) + \gamma^* \mathrm{Cov}(T, \tilde{T}) + \alpha_2^* \mathrm{Cov}(A, \tilde{T}) + \beta_2 \mathrm{Cov}(B, \tilde{T}) + \mathrm{Cov}(\varepsilon_Y^*, \tilde{T}).$$

Because $\mathrm{Cov}(X, \tilde{T}) = 0$, $\mathrm{Cov}(T, \tilde{T}) = \mathrm{Var}(\tilde{T})$, and $\mathrm{Cov}(\varepsilon_Y^*, \tilde{T}) = 0$,

$$\mathrm{Cov}(Y, \tilde{T}) = \gamma^* \mathrm{Var}(\tilde{T}) + \alpha_2^* \mathrm{Cov}(A, \tilde{T}) + \beta_2 \mathrm{Cov}(B, \tilde{T}).$$

5) Compute $\mathrm{Cov}(A, \tilde{T})$ and $\mathrm{Cov}(B, \tilde{T})$, then plug in

Using $\tilde{T} = \alpha_T \tilde{A} + \beta_T \tilde{B} + \varepsilon_T$,

$$\mathrm{Cov}(A, \tilde{T}) = \alpha_T \mathrm{Var}(\tilde{A}) + \beta_T \mathrm{Cov}(A, \tilde{B}),$$

$$\mathrm{Cov}(B, \tilde{T}) = \beta_T \mathrm{Var}(\tilde{B}) + \alpha_T \mathrm{Cov}(B, \tilde{A}),$$

with residualization on X, $\mathrm{Cov}(A, \tilde{B}) = \mathrm{Cov}(B, \tilde{A}) = \mathrm{Cov}(\tilde{A}, \tilde{B})$. Hence,

$$\mathrm{Cov}(Y, \tilde{T}) = \gamma^* \mathrm{Var}(\tilde{T}) + \alpha_T \alpha_2^* \mathrm{Var}(\tilde{A}) + \beta_T \beta_2 \mathrm{Var}(\tilde{B}) + (\alpha_2^* \beta_T + \beta_2 \alpha_T) \mathrm{Cov}(\tilde{A}, \tilde{B}).$$

Therefore,

$$\mathrm{Cov}(Y,\tilde{T}) = (\gamma + \eta\zeta)\mathrm{Var}(\tilde{T}) + (\alpha_2 + \eta\alpha_M)\alpha_T\mathrm{Var}(\tilde{A}) + \beta_2\beta_T\mathrm{Var}(\tilde{B})$$
$$+ [(\alpha_2 + \eta\alpha_M)\beta_T + \beta_2\alpha_T]\mathrm{Cov}(\tilde{A},\tilde{B}).$$

Plugging into the regression anatomy formula gives

$$\boldsymbol{\gamma_{ANCOVA,3C}} = \gamma +$$
$$\boldsymbol{\eta\zeta} + \frac{\boldsymbol{(\alpha_2 + \eta\alpha_M)\alpha_T\mathrm{Var}(\tilde{A}) + \beta_2\beta_T\mathrm{Var}(\tilde{B}) + [(\alpha_2 + \eta\alpha_M)\beta_T + \beta_2\alpha_T]\mathrm{Cov}(\tilde{A},\tilde{B})}}{\boldsymbol{\mathrm{Var}(\tilde{T})}}$$

where the terms added to γ constitute $\boldsymbol{Bias(\gamma_{ANCOVA,3C})_{OLS}}$.

6) Residual variances and covariances

Under standardization,

$\boldsymbol{\mathrm{Var}(\tilde{A}) = 1 - (\alpha_1 + \beta_1\rho)^2}$,

$\boldsymbol{\mathrm{Var}(\tilde{B}) = 1 - (\beta_1 + \alpha_1\rho)^2}$,

$\boldsymbol{\mathrm{Cov}(\tilde{A},\tilde{B}) = \rho - (\alpha_1 + \beta_1\rho)(\beta_1 + \alpha_1\rho)}$,

$\boldsymbol{\mathrm{Var}(\tilde{T}) = 1 - [\delta_T + \alpha_T(\alpha_1 + \beta_1\rho) + \beta_T(\beta_1 + \alpha_1\rho)]^2}$.

7) Unbiasedness conditions

Assuming $\mathrm{Var}(\tilde{T}) > 0$, unbiasedness holds if

$$\eta\zeta = 0 \quad \text{so} \quad \eta = 0 \text{ or } \zeta = 0,$$

and at least one of the following holds:

i. No confounding: $\alpha_T = \beta_T = 0$,

ii. No effect of either confounder on the outcome: $\alpha_2 + \alpha_M\eta = 0$ and $\beta_2 = 0$,

iii. Perfect proxy pretest X: $\mathrm{Var}(\tilde{A}) = \mathrm{Var}(\tilde{B}) = \mathrm{Cov}(\tilde{A},\tilde{B}) = 0$.

Therefore, $Bias(\gamma_{ANCOVA,3C})_G$ captures the core confounding component and the omitted mediator component identified by the graphical approach, whereas $Bias(\gamma_{ANCOVA,3C})_{OLS}$

expresses the total ANCOVA structural bias with the residualization-induced scaling implied by the FWL theorem. Under Scenario 3, Case C, ANCOVA still conditions only on the pretest X, but the data-generating DAG now includes two correlated unobserved confounders A and B, and an unobserved mediator M. The graphical approach therefore implies two distinct sources of structural bias: (a) core noncausal confounding that remains after conditioning on X, now including confounding transmitted through M (so A's contribution enters through its total effect on Y, $\alpha_2 + \alpha_M\eta$, alongside B's β_2, with cross confounding induced by $\rho \equiv \mathrm{Cov}(A, B)$ and by conditioning on the collider X), and (b) an omitted mediator component from the causal path $T \rightarrow M \rightarrow Y$, equal to $\zeta\eta$, because the target causal treatment effect is the direct effect γ while M is omitted from the analytic model. Under standardization, this yields:

$$\boldsymbol{Bias}(\boldsymbol{\gamma_{ANCOVA,3C}})_{\boldsymbol{G}} = \boldsymbol{\zeta\eta} + \boldsymbol{\alpha_T}(\boldsymbol{\alpha_2} + \boldsymbol{\alpha_M\eta}) + \boldsymbol{\beta_T\beta_2} + (\boldsymbol{\alpha_T\beta_2} + \boldsymbol{\beta_T\alpha_2} + \boldsymbol{\beta_T\alpha_M\eta})(\boldsymbol{\rho} + \boldsymbol{\alpha_1\beta_1}).$$

The regression anatomy corroborates this structure: the mediation term $\zeta\eta$ enters additively, whereas the confounding components transmitted through residualization on X and, with two confounders, are scaled by the residual variances and covariance structure $\mathrm{Var}(\tilde{A})$, $\mathrm{Var}(\tilde{B})$, and $\mathrm{Cov}(\tilde{A}, \tilde{B})$. Accordingly, unbiasedness in Case C requires (i) no causal mediation through M ($\zeta\eta = 0$) and (ii) elimination of the remaining confounding, for example by $\alpha_T = \beta_T = 0$ or by $\beta_2 = 0$ together with $\alpha_2 + \alpha_M\eta = 0$. By contrast, "perfect proxy" type restrictions are better viewed not as independent unbiasedness conditions, but as conditions that determine how the OLS residualization scales the core confounding bias.

4.3.3.3. Change Score Analysis: Graphical Approach

Under Scenario 3, Case C, the data generating DAG extends Case B by adding an unobserved mediator M on the treatment to outcome process ($T \rightarrow M \rightarrow Y$) while retaining the two unobserved confounders A and B and their correlation ρ. Because the change score analytic

model remains the simple regression of C on T, it conditions on neither X nor M, so introducing M creates additional open backdoor paths from T to C that run through the segment $M \to Y \to C$. Compared with Case B, this adds nine open backdoor paths, increasing the total number from 18 to 27:

Confounding via the pretest X

1) $T \leftarrow X \to C$: $-\delta_T$,

2) $T \leftarrow X \to Y \to C$: $\delta_T \tau$,

3) $T \leftarrow X \to M \to Y \to C$: $\delta_T \delta_M \eta$.

Confounding via A

4) $T \leftarrow A \to Y \to C$: $\alpha_T \alpha_2$,

5) $T \leftarrow A \to X \to C$: $-\alpha_T \alpha_1$,

6) $T \leftarrow A \to X \to Y \to C$: $\alpha_T \alpha_1 \tau$,

7) $T \leftarrow A \to M \to Y \to C$: $\alpha_T \alpha_M \eta$,

8) $T \leftarrow A \to X \to M \to Y \to C$: $\alpha_T \alpha_1 \delta_M \eta$.

Confounding via B

9) $T \leftarrow B \to Y \to C$: $\beta_T \beta_2$,

10) $T \leftarrow B \to X \to C$: $-\beta_T \beta_1$,

11) $T \leftarrow B \to X \to Y \to C$: $\beta_T \beta_1 \tau$,

12) $T \leftarrow B \to X \to M \to Y \to C$: $\beta_T \beta_1 \delta_M \eta$.

Confounding via the A–B **correlation (**$A \leftarrow U \to B$ or $B \leftarrow U \to A$**)**

Under standardization, $\mathrm{Cov}(A, B) = \rho_A \rho_B = \rho$.

13) $T \leftarrow A \leftarrow U \to B \to Y \to C$: $\alpha_T \rho \beta_2$,

14) $T \leftarrow A \leftarrow U \to B \to X \to C$: $-\alpha_T \rho \beta_1$,

15) $T \leftarrow A \leftarrow U \rightarrow B \rightarrow X \rightarrow Y \rightarrow C$: $\alpha_T \rho \beta_1 \tau$,

16) $T \leftarrow A \leftarrow U \rightarrow B \rightarrow X \rightarrow M \rightarrow Y \rightarrow C$: $\alpha_T \rho \beta_1 \delta_M \eta$,

17) $T \leftarrow B \leftarrow U \rightarrow A \rightarrow Y \rightarrow C$: $\beta_T \rho \alpha_2$,

18) $T \leftarrow B \leftarrow U \rightarrow A \rightarrow X \rightarrow C$: $-\beta_T \rho \alpha_1$,

19) $T \leftarrow B \leftarrow U \rightarrow A \rightarrow X \rightarrow Y \rightarrow C$: $\beta_T \rho \alpha_1 \tau$,

20) $T \leftarrow B \leftarrow U \rightarrow A \rightarrow M \rightarrow Y \rightarrow C$: $\beta_T \rho \alpha_M \eta$,

21) $T \leftarrow B \leftarrow U \rightarrow A \rightarrow X \rightarrow M \rightarrow Y \rightarrow C$: $\beta_T \rho \alpha_1 \delta_M \eta$.

Confounding via X and the two confounders

22) $T \leftarrow X \leftarrow A \rightarrow Y \rightarrow C$: $\delta_T \alpha_1 \alpha_2$,

23) $T \leftarrow X \leftarrow B \rightarrow Y \rightarrow C$: $\delta_T \beta_1 \beta_2$,

24) $T \leftarrow X \leftarrow A \leftarrow U \rightarrow B \rightarrow Y \rightarrow C$: $\delta_T \alpha_1 \rho \beta_2$,

25) $T \leftarrow X \leftarrow B \leftarrow U \rightarrow A \rightarrow Y \rightarrow C$: $\delta_T \beta_1 \rho \alpha_2$,

26) $T \leftarrow X \leftarrow A \rightarrow M \rightarrow Y \rightarrow C$: $\delta_T \alpha_1 \alpha_M \eta$,

27) $T \leftarrow X \leftarrow B \leftarrow U \rightarrow A \rightarrow M \rightarrow Y \rightarrow C$: $\delta_T \beta_1 \rho \alpha_M \eta$.

In addition to these noncausal confounding components, the fully causal mediation route $T \rightarrow M \rightarrow Y \rightarrow C$ contributes an omitted mediator bias term:

$$T \rightarrow M \rightarrow Y \rightarrow C: \zeta \eta.$$

Summing the path-tracing products over the number of 27 open backdoor paths and adding the mediation term yields the structural bias:

$$\boldsymbol{Bias(\gamma_{Change,3C})_G} =$$

$$\boldsymbol{\zeta\eta + \delta_T[\alpha_1\alpha_2 + \beta_1\beta_2 - 1 + \rho(\alpha_1\beta_2 + \alpha_2\beta_1) + \tau + \eta\{\alpha_M(\alpha_1 + \rho\beta_1) + \delta_M\}]}$$

$$\boldsymbol{+\alpha_T[(\alpha_2 - \alpha_1) + \rho(\beta_2 - \beta_1) + \tau(\alpha_1 + \rho\beta_1) + \eta\{\alpha_M + \delta_M(\alpha_1 + \rho\beta_1)\}]}$$

$$\boldsymbol{+ \beta_T[(\beta_2 - \beta_1) + \rho(\alpha_2 - \alpha_1) + \tau(\beta_1 + \rho\alpha_1) + \eta\{\alpha_M\rho + \delta_M(\beta_1 + \rho\alpha_1)\}]}.$$

4.3.3.4. Change Score Analysis: OLS

Under Scenario 3, Case C, the mediator and outcome equations are $M = \alpha_M A + \delta_M X + \zeta T + \varepsilon_M$ and $Y = \alpha_2 A + \beta_2 B + \tau X + \gamma T + \eta M + \varepsilon_Y$. Substituting M into Y yields:

$$Y = (\tau + \eta\delta_M)X + (\gamma + \eta\zeta)T + (\alpha_2 + \eta\alpha_M)A + \beta_2 B + (\varepsilon_Y + \eta\varepsilon_M).$$

Define $\tau^\star = \tau + \eta\delta_M$, $\gamma^\star = \gamma + \eta\zeta$, and $\alpha_2^\star = \alpha_2 + \eta\alpha_M$ so that $Y = \tau^\star X + \gamma^\star T + \alpha_2^\star A + \beta_2 B + \varepsilon_Y^\star$.

Then

$$\mathrm{Cov}(Y,T) = \tau^\star \mathrm{Cov}(X,T) + \gamma^\star \mathrm{Var}(T) + \alpha_2^\star \mathrm{Cov}(A,T) + \beta_2 \mathrm{Cov}(B,T).$$

Therefore,

$$\boldsymbol{\gamma_{Change,3C} = \gamma + \eta\zeta + \delta_T C_\delta + \alpha_T C_\alpha + \beta_T C_\beta}$$

where the terms added to γ constitute $\boldsymbol{Bias(\gamma_{Change,3C})_{OLS}}$, and the bracket terms are:

$$\boldsymbol{C_\delta = \alpha_1\alpha_2 + \beta_1\beta_2 - 1 + \rho(\alpha_1\beta_2 + \alpha_2\beta_1) + \tau + \eta\{\alpha_M(\alpha_1 + \rho\beta_1) + \delta_M\}},$$

$$\boldsymbol{C_\alpha = (\alpha_2 - \alpha_1) + \rho(\beta_2 - \beta_1) + \tau(\alpha_1 + \rho\beta_1) + \eta\{\alpha_M + \delta_M(\alpha_1 + \rho\beta_1)\}},$$

$$\boldsymbol{C_\beta = (\beta_2 - \beta_1) + \rho(\alpha_2 - \alpha_1) + \tau(\beta_1 + \rho\alpha_1) + \eta\{\alpha_M\rho + \delta_M(\beta_1 + \rho\alpha_1)\}}.$$

Therefore, $Bias(\gamma_{Change,3C})_G$ remains identical to $Bias(\gamma_{Change,3C})_{OLS}$. A practically interpretable sufficient unbiasedness conditions begins by requiring $\zeta\eta = 0$, meaning either $\zeta = 0$ or $\eta = 0$ and thus no causal mediation from T to Y through M. It also requires $\delta_T = 0$, meaning there is no dynamic causal relationship between the pretest and treatment assignment. Under $\delta_T = 0$, unbiasedness further requires the common trend (time invariant confounding) conditions that incorporate both the A to B correlation ρ and the mediator related paths:

$$\alpha_2 + \eta\,[\alpha_M + (\tau + \eta\delta_M)\alpha_1] + \rho\,[\beta_2 + (\tau + \eta\delta_M)\beta_1] = \alpha_1 + \rho\beta_1, and$$

$$\beta_2 + (\tau + \eta\delta_M)\beta_1 + \rho\,[\alpha_2 + \eta\,(\alpha_M + (\tau + \eta\delta_M)\alpha_1)] = \beta_1 + \rho\alpha_1.$$

Substantively, these conditions state that, once the A to B association is taken into account, each confounder's total effect on the untreated posttest (including paths operating through the mediator and the pretest to posttest carryover structure) matches its effect on the pretest. A simplified but stronger version sets $\alpha_2 = \alpha_1$ and $\beta_2 = \beta_1$, and additionally imposes $\tau = 0$ and $\eta = 0$. Alternatively, unbiasedness also holds if $\alpha_T = 0$and $\beta_T = 0$, so neither unobserved confounder affects treatment assignment. In addition, unbiasedness is also technically possible through exact bias offsetting, which occurs when $\eta\zeta + \delta_T C_\delta + \alpha_T C_\alpha + \beta_T C_\beta = 0$ (with C_δ, C_α, and C_β defined above).

5. Discussion and Conclusion

This dissertation started from a practical but unresolved gap in the two wave nonequivalent control group literature: even when researchers agree that unobserved variables can create structural bias, it is not always clear how to characterize that bias in a way that is both causally explicit in terms of a DAG and algebraically quantified by the OLS for the analytic method being used. Two representative studies illustrate this disconnect. Kim and Steiner (2021) use causal DAGs with the backdoor criterion and path-tracing to derive and interpret bias for change score analysis, whereas Ludtke and Robitzsch (2025) derive bias comparisons between ANCOVA and change score analysis using the OLS based population formulas without an explicit DAG based account. What was missing is a single framework that places the graphical approach and the OLS results side by side, shows when they coincide and when they differ, and explains why those agreements and differences arise for both change score analysis and ANCOVA within the same set of two wave data generating scenarios.

5.1. Central Findings

The findings presented in chapter 4 directly fill that gap. Across all scenarios and cases, the graphical approach consistently clarifies open noncausal paths remain once an analytic model is chosen, and therefore clarifies the causal assumptions needed for an unbiased treatment effect. On the algebraic side, the OLS results mirror these graphical conclusions. Specifically, in change score analysis, the relevant estimator is a simple regression coefficient, so the bias can be derived directly from the standard simple regression coefficient formula for the change score on treatment. Second, for ANCOVA, the estimator is a partial regression coefficient after conditioning on the pretest, so the OLS results rely on the Frisch–Waugh–Lovell theorem to make explicit how the

remaining core confounding component identified by the graph enters the ANCOVA coefficient and how it is rescaled by a residualization induced variance and covariance terms. Given this, several central findings emerged from this study.

The first central finding is that, for ANCOVA, the graphical approach identifies the open backdoor confounding paths in the DAG and expresses the resulting core bias via path-tracing, whereas quantifying the exact structural bias magnitude additionally requires the residual variance (and, in case of two confounders, residual covariance) structure implied by the FWL theorem. Because the ANCOVA estimator is a partial regression coefficient, using DAGs is most informative about the post conditioning causal structure: it shows which backdoor paths remain open after adjustment for X and therefore yields a path-tracing expression for the resulting core noncausal confounding component. What the graphical approach does not display directly is the rescaling created by ANCOVA's partial regression coefficient definition. The FWL theorem makes this additional step explicit by expressing the ANCOVA coefficient in terms of residualized treatment variation after regressing Ton X, which implies that the total ANCOVA structural bias equals the core confounding component multiplied by a residualization-induced scaling term. In the one confounder setting this scaling is governed by the relevant residual variances, and with two confounders it also depends on the residual covariance structure after partialling out X. Expressing this rescaling explicitly in terms of residual variances or residual covariances makes the mechanism transparent: ANCOVA rescales the core confounding component according to how much confounder variation remains after partialling out X relative to how much treatment variation remains after partialling out X.

The second central finding is that the graphical approach alone is still enough to state unbiasedness conditions for ANCOVA even though it does not produce the residual variance

scaling terms. Across scenarios, the unbiasedness conditions that are most interpretable are those that correspond to removing the remaining open backdoor paths from treatment to posttest. These are assumption of "no causal effect of the unobserved confounder on treatment" or "no causal effect of the unobserved confounder on the posttest." In contrast, the "perfect proxy" type conditions (for example, conditions that force residual variation in an unobserved confounder to collapse to zero after partialing out the pretest) are best viewed not as realistic assumptions one would lean on for unbiasedness, but as limiting cases in which the residualization step leaves essentially no confounder variation to be rescaled.

The third central finding is that, under the dissertation's direct effect definition of the treatment effect, the graphical approach makes the omitted mediator component of structural bias explicit in the mediator cases (Case C). Importantly, this omitted mediation component is not created by the definition alone; it arises only when the data generating DAG includes a causal path $T \to M \to Y$ that is omitted from the analytic model and it is zero when the $T \to M$ edge is absent, $\zeta = 0$, so that only $X \to M \to Y$ remains. In Case C, the data generating model therefore contains both a direct path $T \to Y$ and an indirect path $T \to M \to Y$, while the analytic models omit M. Under this setup, structural bias has two distinct sources: a confounding component from any open noncausal backdoor paths and an omitted mediation component from the causal $T \to M \to Y$ path that is excluded from the analytic model. The graphical approach makes this separation easy to see by distinguishing the remaining open backdoor structure from the omitted causal mediation path. The OLS results confirm the same structure algebraically. The omitted mediation component enters as a separate additive term corresponding to the causal path $T \to M \to Y$, and it does not carry ANCOVA's residualization induced scaling because the T driven part of $\mathrm{Cov}(\tilde{T}, M)$ (equivalently, $\mathrm{Cov}(\tilde{T}, \widetilde{M})$) is proportional to $\mathrm{Var}(\tilde{T})$ and therefore cancels with the

$\mathrm{Var}(\tilde{T})$ denominator in the ANCOVA coefficient. By contrast, the confounding component enters through ratios such as $\mathrm{Cov}(\tilde{T}, \tilde{A})/\mathrm{Var}(\tilde{T})$ (and, with multiple confounders, the residual covariance structure) and thus follows ANCOVA's residualization based scaling logic. Consistent with the omitted mediation derivations, when the $T \to M$ edge is absent ($\zeta = 0$), so that only an $X \to M \to Y$ path remains, the omitted mediation component is zero because there is no causal effect of T transmitted through M.

The fourth and final central finding is that, for change score analysis, the graphical approach and the OLS results yield identical structural bias expressions. Because the change score estimator is a simple regression coefficient without residualization, the structural bias from open-backdoor path-tracing coincides exactly with the OLS bias expression across scenarios: the bias is just the core confounding component implied by the remaining open backdoor structure, without an additional residualization-induced scaling term. Across these results, unbiasedness repeatedly hinges on a generalized common trend requirement, that for each unobserved confounder, its total causal effect on the untreated posttest is the same as its effect on the pretest, together with the no dynamic causal relationship between the pretest and treatment assignment condition. An alternative sufficient route to unbiasedness is also consistent throughout, that treatment assignment is not caused by the observed pretest and not caused by any unobserved confounders (i.e., treatment exogeneity; Hernán & Robins, 2020, p. 14), but this is rarely plausible in nonequivalent control group designs.

5.2. Methodological Contributions

Based on these central findings, this dissertation's methodological contribution is the establishment of explicit link between the tracing based graphical approach and the OLS estimation for structural bias. Wright (1921) originally developed path analysis to relate observed

correlation or covariance patterns to a hypothesized system of functional relations by decomposing associations into sums of product of path coefficients along valid paths in a path diagram. Pearl (2013) returns to Wright's tracing rules and uses them to compute covariances from *d*-connected paths, and then shows that once those covariances are available, standard partial regression formulas can be used to obtain conditional regression quantities and to measure biases created by conditioning. Building on this foundation, Kim and Steiner (2021) applied path-tracing to two wave change score graphs by explicitly listing open noncausal backdoor paths and expressing the resulting noncausal association as products of structural coefficients, thereby obtaining bias expressions and clarifying when bias can be removed by offsetting rather than blocking. The present contribution extends this tracing based bias analysis from change score analysis to ANCOVA, where the treatment effect estimator is not a simple regression coefficient but a partial regression coefficient after conditioning on the pretest. In this setting, the graphical approach continues to be highly informative because it identifies the remaining open noncausal confounding component, the core backdoor bias that persists even after conditioning choices are made. Importantly, the graphical approach alone can still determine when ANCOVA yields an unbiased treatment effect, because those conditions correspond to removing the remaining open backdoor paths

For ANCOVA, the graphical approach is sufficient to state the causal assumptions needed for an unbiased treatment effect, but quantifying the exact structural bias magnitude requires an additional step because the estimator is a partial regression coefficient defined after residualizing treatment with respect to the pretest. In the two-wave setting, this dissertation shows systematically how the graphical approach and the Frisch Waugh Lovell regression anatomy complement each other for ANCOVA. The graphical approach makes the causal assumptions and the remaining

open backdoor structure under ANCOVA easy to see, and it isolates the core confounding component that persists after conditioning on X. The FWL theorem's residualization step then makes explicit what the graph alone does not display: because ANCOVA uses $\tilde{T}$, the residualized treatment after regressing T on X, the remaining core confounding component enters the ANCOVA coefficient through a residualization induced scaling factor that depends on residual variances and, with two confounders, residual covariances after partialling out X. Importantly, when the goal is to quantify the exact bias magnitude, regression anatomy offers a simpler and more transparent route than computing every implied covariance and plugging them into a generic partial regression formula. Instead, we can take the core confounding component shown by the DAG and then apply a single scaling step based on the relevant residual variance or residual covariance quantities. In this way, the dissertation demonstrates that the graphical approach can be used to visualize why ANCOVA bias arises, while the residualization step provides an accessible bridge to the exact magnitude when needed, making the overall framework easier to understand and use for applied social science researchers.

This dissertation shows how ANCOVA can amplify structural bias when unobserved confounding remains after conditioning on the pretest. Prior discussions of Lord's paradox and related debates often emphasize the possibility of bias amplification under ANCOVA (Pearl, 2011; Steiner & Kim, 2016), typically by focusing on the idea that conditioning can shrink the denominator of the bias terms when the conditioning pretest causally affects treatment assignment. In that framing, if the pretest does not remove all confounding, but it does strongly cause treatment assignment, then the remaining bias is evaluated relative to a smaller residualized treatment variance, so the remaining bias can be inflated. The present results show that this denominator-focused amplification logic can be stated more directly in terms of how residualization rescales

the core confounding bias identified by the graphical approach. Under the OLS results using the FWL theorem, ANCOVA's structural bias can be written as a product of two pieces: (1) a core confounding component that corresponds to the remaining open backdoor structure after conditioning on the pretest, and (2) a residualization-induced scaling term due to defining the ANCOVA treatment coefficient from the residual variance T after regressing T on X. This separation captures the two forces emphasized in the amplification literature: conditioning on X may reduce how much confounding remains, but if T is largely explained by X, the remaining confounding can have a larger impact on the ANCOVA coefficient because the residualized treatment variation is small.

Under the one unobserved confounder setting (Scenario 2), this point is especially transparent. After placing A and T on a common scale (e.g., via standardization), the ANCOVA bias can be written as the core confounding component multiplied by a scaling term of the form $\mathrm{Var}(\tilde{A})/\mathrm{Var}(\tilde{T})$, where $\tilde{A}$ is the part of the unobserved confounder not explained by X and $\tilde{T}$ is the part of treatment assignment not explained by X. This makes the amplification condition explicit: holding the core confounding relationship fixed, ANCOVA magnifies that component when $\mathrm{Var}(\tilde{A}) > \mathrm{Var}(\tilde{T})$. This criterion also aligns with the kind of scenario often used to motivate amplification concerns in nonequivalent designs. Amplification is most plausible in Scenario 2, Case B honors-track placement setting when treatment assignment is driven primarily by the observed pretest X, yet X is an imperfect proxy for the latent ability confounder A under a mixture perspective. Here, A represents stable math ability, while X reflects ability only partially and incorporates other stock-like and flow-like components such as test-taking proficiency, prior test preparation, situational motivation, and other time-specific factors. As a result, conditioning on X does not eliminate much of the latent ability variation, so meaningful within-X differences in

Acan remain. At the same time, honors placement $T$is often tied closely to baseline test scores (e.g., via a cutoff or a rule that heavily weights X), which leaves relatively little remaining variation in Tafter conditioning on X. Under these conditions, $\mathrm{Var}(\tilde{A})$ can remain substantial while $\mathrm{Var}(\tilde{T})$ becomes small, pushing the scaling term above one. For instance, if $\mathrm{Var}(\tilde{A}) = 0.40$ and $\mathrm{Var}(\tilde{T}) = 0.10$, then $\mathrm{Var}(\tilde{A})/\mathrm{Var}(\tilde{T}) = 4.0$, implying that the remaining confounding component is multiplied by 4.0 in the ANCOVA bias.

5.3. Practical Implications for Applied Research

This dissertation has several implications for applied researchers in Education, Psychology, and related fields who must choose between ANCOVA and change-score analysis in two-wave nonequivalent control-group designs. The central practical issue is not merely that the two methods can produce different treatment-effect estimates, but why they differ and what that difference implies about the causal assumptions required for each method to recover the population treatment effect.

A useful way to translate these results into practice is to treat the graphical approach as a decision framework. Rather than selecting ANCOVA or change-score analysis as routine methods for "controlling baseline differences," researchers can ask a more informative set of questions: Is treatment assignment related to pretest performance or to variables closely associated with the pretest? Are the major confounders observed and measurable? Does the pretest adequately capture the latent differences that influence both treatment assignment and outcomes? If important confounders are unobserved, are the common-trend assumption and the relevant restrictions on dynamic causal relationships involving the pretest substantively credible? Framing the problem in this way allows researchers to justify analytic choices in causal terms that are often easier to communicate than covariance algebra alone.

When the major confounders can be observed and controlled, ANCOVA is generally the preferred method. This is especially true when treatment assignment is related to baseline performance or related baseline characteristics, because conditioning on the pretest blocks an important class of selection-related backdoor paths. For example, if students are assigned to a remedial intervention because of low baseline achievement, ANCOVA will often be more defensible than change-score analysis, provided that the pretest and other baseline covariates adequately represent the selection process. At the same time, the results of this dissertation clarify that ANCOVA is not automatically unbiased. If the pretest is only a weak proxy for the latent factors that jointly influence treatment assignment and posttest outcomes, residual confounding may remain after conditioning, and the remaining bias may be rescaled. In practical terms, this means that researchers should not treat baseline adjustment as a mechanical solution. Instead, they should explain why the observed pretest and covariates are plausible proxies for the relevant confounding process and, when possible, improve the design by measuring additional baseline covariates.

Change-score analysis also remains useful, but under stronger assumptions. Its practical advantage is that it can recover an unbiased treatment effect without conditioning on the pretest when the common-trend assumption and the relevant restrictions on dynamic causal relationships involving the pretest hold. In other words, if researchers have strong substantive reasons to believe that unobserved group differences are sufficiently time-invariant and that the pretest does not generate problematic dynamic bias, then change-score analysis can be a reasonable analytic choice. For example, if two nonrandomized groups differ in stable baseline characteristics but would otherwise have changed in parallel in the absence of treatment, change-score analysis may be viable even when some confounding remains unobserved. This point is important because it

clarifies that change-score analysis should not be dismissed simply because it leaves the pretest unconditioned on. Rather, it is a useful method when its identifying assumptions can be defended. In this sense, the emphasis in Kim and Steiner on common trends and bias offsetting remains substantively meaningful, but the conditions that justify those arguments should be made explicit in applied work rather than taken for granted.

Taken together, these results suggest a practical decision rule for applied research. If researchers can measure and control the main confounders, especially when treatment assignment is related to baseline performance, ANCOVA should usually be the primary analysis. If substantial unobserved confounding remains a concern but the common-trend assumption and the relevant restrictions on dynamic causal relationships involving the pretest are credible, change-score analysis may be a defensible alternative. In ambiguous settings, the most transparent practice is to present a simple DAG, state which backdoor paths each method blocks or leaves open, justify the preferred analytic model in terms of those assumptions, and, when feasible, report the alternative estimate as a robustness or sensitivity analysis. This kind of translation is especially useful for practitioners, because it converts an abstract paradox into a concrete set of design and modeling decisions.

5.4. Limitations and Future Directions

This dissertation uses diverse two-wave scenarios to compare structural bias under ANCOVA and change score analysis using the graphical approach and the OLS results. This setup makes the graphical and the OLS bias directly comparable across analytic methods in a more straightforward way, but it also leaves several limitations that motivate future directions.

The first limitation is that the application of the graphical approach is restricted to two-wave, single-level scenarios. The scenarios were intentionally limited to a simple two-wave

structure to keep the causal logic transparent and to make ANCOVA and change score comparisons clean. Many applied evaluations, however, involve more complex designs, including three or more waves, treatment timing that varies across units, and multilevel structures such as students nested within classrooms and schools. These features can introduce additional sources of bias and additional open paths that are not visible in a two-wave, single-level setup, and they can also change what "conditioning on the pretest" means when baseline information exists at multiple levels. A natural next step is therefore to extend the same graphical approach and its OLS using FWL theorem to multi-wave settings and multilevel settings, where residualization and open-path logic can be examined under realistic longitudinal and clustered data structures.

The second limitation is that the pretest as a mixture perspective of stocks and flows is used as a theoretical organizing device in this dissertation rather than as an empirically tested claim. The cases were organized by treating the pretest as a mixture of stock-like components and flow-like components that can carry forward, which helps motivate why the direct or indirect causal effect of the pretest on posttest may or may not be plausible. However, the mixture perspective itself does not uniquely identify which stock-like, and flow-like components are present in any given application or how much each contributes to the observed pretest, and these components' proportions can hardly be tested cleanly with only two-waves. Relatedly, the pretest as a mixture perspective draws conceptually on the idea of an "accumulating factor" discussed in Usami, Murayama, and Hamaker (2019), but the relationship between that accumulating factor and the pretest as a mixture perspective is not specified in a one-to-one way. In their longitudinal SEM framing, an accumulating factor refers to a time-invariant latent factor whose influence on later waves can grow because it operates not only through direct effects but also through indirect effects carried by lagged relations across waves. In contrast, the pretest as a mixture perspective is a

conceptual account of why the observed pretest score may combine durable components (i.e., stocks) and time-specific components (i.e., flows). With only two waves, it is difficult to empirically separate stable and time specific components, or to test whether persistence in the pretest to posttest relation reflects stable trait like carryover, accumulation of stable influences through the lagged structure, or accumulation of initially time specific processes that become reinforced over time because two time points provide too little information to distinguish these sources of persistence from the observed pretest posttest covariance pattern. A clear future direction is therefore to test and refine the pretest as a mixture perspective in multi-wave designs using longitudinal SEM or panel models that can separate stable and time specific components and evaluate whether pretest to posttest carryover reflects stable persistence, accumulation through lagged dynamics, or reinforcement and habit formation processes that allow time specific components to persist across waves.

5.5. Conclusion

This dissertation addressed a gap in the literature on two-wave nonequivalent control-group designs. Although researchers have long recognized that unobserved confounding can bias treatment-effect estimates, it has remained unclear how to characterize that bias in a way that is both causally explicit and algebraically exact for the specific analytic method being used. To address this problem, the study pursued two goals. First, it compared the structural bias and sufficient unbiasedness conditions of ANCOVA and change-score analysis across a set of two-wave data-generating scenarios. Second, it extended the graphical approach so that it could fully characterize the structural bias of the population-level ANCOVA estimator by linking directed acyclic graph (DAG) path tracing to population-level ordinary least squares (OLS) through the Frisch–Waugh–Lovell theorem.

The results showed that the relationship between the graphical and OLS characterizations differed by analytic method. For change-score analysis, the graphical approach reproduced the exact OLS structural bias expressions across the scenarios. Across these results, unbiasedness depended primarily on a common-trend condition together with the required restrictions on dynamic causal relationships involving the pretest. For ANCOVA, the graphical approach identified the remaining open backdoor confounding after conditioning on the pretest and therefore recovered the core confounding component of structural bias. Because the ANCOVA estimator is a partial regression coefficient, however, the graph alone did not determine the total magnitude of that bias. The OLS characterization using the Frisch–Waugh–Lovell theorem supplied the missing step by showing that total ANCOVA structural bias equals the remaining core confounding component multiplied by a residualization-induced scaling term.

Taken together, these findings show that DAG-based reasoning and population-level OLS provide complementary rather than competing characterizations of structural bias. The graphical approach identifies the remaining open backdoor structure and the assumptions required for unbiasedness, whereas the OLS characterization makes explicit the variance- and covariance-based scaling needed to quantify total bias when conditioning is involved. By placing these results within a common set of two-wave scenarios, this dissertation clarifies when ANCOVA and change-score analysis coincide, when they differ, and how structural bias in two-wave nonequivalent control-group designs can be characterized in a way that is both causally explicit and algebraically exact. In this sense, the dissertation's principal methodological contribution is to establish an explicit link between DAG-based path tracing and population-level OLS, thereby providing a more accessible basis for reasoning about structural bias than OLS coefficient formulas alone.

References

Allison, P. D. (1990). Change score as dependent variables in regression analysis. *Sociological Methodology, 20,* 93-114.

Angrist, J. D., & Pischke, J. S. (2009). *Mostly Harmless Econometrics: An empiricist's companion.* Princeton University Press.

Bollen, K. A. (1989). *Structural equations with latent variables*. John Wiley & Sons.

Campbell, D. T., & Erlebacher, A. (1970). How regression artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful. In: J. Hellmuth (Ed.), *The Disadvantaged Child Vol. III: Compensatory Education: A National Debate*. New York: Brunner/Mazel.

Campbell, D. T., & Kenny, D. A. (1999). *A primer on regression artifacts*. The Guilford Press.

Cicirelli, V. G. (1969). *The Impact of Head Start: An Evaluation of the Effects of Head Start on Children's Cognitive and Affective Development*, Westinghouse Learning Corp., and Ohio University under contract B89-4536. Washington, D.C: Office of Economic Opportunity.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Routledge.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Rand Mc-Nally.

Cunningham, S. (2021). *Causal inference: The mixtape*. Yale University Press.

Fisher, R. A. (1970). *Statistical methods for research workers* (14th ed.). Hafner.

Frisch, R., & Waugh, F. V. (1933). Partial time regressions as compared with individual trends. *Econometrica: Journal of the Econometric Society*, 387-401.

Goldberger, A. S. (1962). Best linear unbiased prediction in the generalized linear regression model. *Journal of the American Statistical Association*, *57*(298), 369-375.

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

Holland, P. W., & Rubin, D. B. (1983). On Lord's paradox. In H. Wainer & S. Messick (Eds.), *Principles of modem psychological measurement* (pp. 3-25). Erlbaum.

Hong, G. (2015). *Causality in a social world: Moderation, mediation and spill-over*. John Wiley & Sons.

Imai, K., & Kim, I. S. (2019). When should we use unit fixed effects regression models for causal inference with longitudinal data?. *American Journal of Political Science*, *63*(2), 467-490.

Kaufman, J. H., Covelli, L., & Holmes, P. (2024). Elementary and Middle School Opportunity Structures That Factor into Students' Math Learning: Findings from the American Mathematics Educator Study. Research Report. RR-A2836-2. *RAND Corporation*.

Kenny, D. A. (1975). A quasi-experimental approach to assessing treatment effects in the nonequivalent control group design. *Psychological Bulletin*, *82*(3), 345-362.

Kenny, D. A. (1979). *Correlation and causality.* Wiley.

Kim, Y., & Steiner, P. M. (2021). Gain scores revisited: A graphical models perspective. *Sociological Methods & Research*, *50*(3), 1353-1375.

Lüdtke, O., & Robitzsch, A. (2025). ANCOVA versus change score for the analysis of two-wave data. *The Journal of Experimental Education*, *93*(2), 363-395.

Lord, F. M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin*, *68*(5), 304-305.

Lord, F. M. (1969). Statistical adjustments when comparing preexisting groups. *Psychological Bulletin*, *72*(5), 336.

Lord, F. M. (1975). Lord's Paradox. In S. B. Anderson, S. Ball, & R. T. Murphy (Eds.), *Encyclopedia of Educational Evaluation* (pp. 232-236). Jasey-Bass.

Lovell, M. C. (1963). Seasonal adjustment of economic time series and multiple regression analysis. *Journal of the American Statistical Association*, *58*(304), 993-1010.

Lechner, M. (2010). The estimation of causal effects by difference-in-difference methods. *Foundations and Trends® in Econometrics*, *4*(3), 165-224.

Maxwell, S. E., Delaney, H. D., & Kelley, K. (2018). *Designing experiments and analyzing data: a model comparison perspective* (3rd ed.). Routledge.

Mill, J. S. (2023). *A system of logic, ratiocinative and inductive*. Global Grey eBooks. https://www.globalgreyebooks.com/system-of-logic-ratiocinative-and-inductive-ebook.html (Original work published in 1843)

Morgan, S. L., & Winship, C. (2014). *Counterfactuals and Causal Inference: Methods and Principles for Social Research* (Second edition). Cambridge University Press.

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

Pearl, J. (2011). Invited commentary: understanding bias amplification. *American Journal of Epidemiology*, *174*(11), 1223-1227.

Pearl, J. (2013). Linear models: A useful "microscope" for causal analysis. *Journal of Causal Inference*, *1*(1), 155-170.

Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal inference in statistics: A primer.* Wiley.

Pearl, J. (2016). Lord's paradox revisited – (oh Lord! Kumbaya!). *Journal of Causal Inference*, *4*(2), 20160021.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research: explanation and prediction* (3rd ed.). Harcourt Brace College Publishers.

Rubin, D. B. (1975). Bayesian inference for causality: The importance of randomization. In *Proceedings of the Social Statistics Section of the American Statistical Association* (pp. 233-239). Alexandria, VA: American Statistical Association.

Rubin, D. B. (2008). For objective causal inference, design trumps analysis. *Annals of Applied Statistics, 2*(3), 808-840.

Seber, G. A. F., & Lee, A. J. (2012). *Linear regression analysis* (2nd ed.). John Wiley & Sons.

Shadish, W. R., Cook, T. D., Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*, Wadsworth Cengage Learning.

Stapleton, J. H. (2009). *Linear statistical models* (2nd ed.). Wiley.

Steiner, P. M., & Kim, Y. (2016). The mechanics of omitted variable bias: Bias amplification and cancellation of offsetting biases. *Journal of Causal Inference*, *4*(2), 20160009.

Usami, S., Murayama, K., & Hamaker, E. L. (2019). A unified framework of longitudinal models to examine reciprocal relations. *Psychological Methods, 24*(5), 637–657.

Wainer, H., & Brown, L. M. (2004). Two statistical paradoxes in the interpretation of group differences: Illustrated with medical school admission and licensing data. *The American Statistician*, *58*(2), 117-123.

West, S. G., Biesanz, J. C., & Pitts, S. C. (2000). Causal inference and generalization in field settings: Experimental and quasi-experimental designs. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 40–84). Cambridge University Press.

Whitley, B. E., & Kite, M. E. (2013). *Principles of research in behavioral science* (3rd ed.). Routledge.

Wooldridge, J. M. (2020). *Introductory econometrics: a modern approach* (7th ed.). Cengage.

Wright, D. B. (2006). Comparing groups in a before–after design: When t-test and ANCOVA produce different results. *British Journal of Educational Psychology, 76*(3)*,* 663–675.

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

Venter, A., Maxwell, S. E., & Bolig, E. (2002). Power in randomized group comparisons: the value of adding a single intermediate time point to a traditional pretest-posttest design. *Psychological Methods*, *7*(2), 194-209.