HPOG
Health Profession Opportunity Grants
OFFICE OF FAMILY ASSISTANCE

Using Within-Site Experimental Evidence to Reduce Cross-Site Attributional Bias in Connecting Program Components to Program Impacts

OPRE Report 2017-13

January 2017

Using Within-Site Experimental Evidence to Reduce Cross-Site Attributional Bias in Connecting Program Components to Program Impacts

Special Topic Paper

OPRE Report 2017-13

January 2017

Stephen H. Bell, Eleanor L. Harvill, Shawn R. Moulton and Laura Peck, Abt Associates,
4550 Montgomery Avenue, Suite 800 North, Bethesda, MD 20814

Submitted to:
Hilary Forster and Nicole Constance, Contracting Officer's Technical Representatives
Office of Planning, Research and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services

Contract No. HHSP23320095624WC, Task Order HHSP23337012T

Project Director: Gretchen Locke
Abt Associates Inc.
55 Wheeler Street
Cambridge, MA 02138

Suggested citation: Stephen H. Bell, Eleanor L. Harvill, Shawn R. Moulton, and Laura Peck. (2017). *Using Within-Site Experimental Evidence to Reduce Cross-Site Attributional Bias in Connecting Program Components to Program Impacts,* OPRE Report #2017-13. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Disclaimer: The views expressed in this publication do not necessarily reflect the views or policies of the Office of Planning, Research and Evaluation, the Administration for Children and Families, or the U.S. Department of Health and Human Services. This report and other reports sponsored by the Office of Planning, Research and Evaluation are available at www.acf.hhs.gov/opre.

Acknowledgements: This paper was supported in part by a grant from the Open Society Foundations and by the HPOG Impact Study, funded by the U.S. Department of Health and Human Services, for presentation at the 2015 Annual Fall Research Conference of the Association for Public Policy Analysis and Management (APPAM) in Miami, Florida. We are grateful for research and editorial assistance from Edward Bein and Sarah Sahni, for critical input from Rob Olsen, and for editorial assistance from Bry Pollack. We also recognize feedback from participants in Abt's Journal Author Support Group, as well as from APPAM conference panelists and discussant Jeff Smith (University of Michigan).

Overview

This paper considers a new method, called Cross-Site Attributional Model Improved by Calibration to Within-Site Individual Randomization Findings (CAMIC), which seeks to reduce bias in analyses that researchers use to understand what about a program's structure and implementation leads its impact to vary.

Randomized experiments—in which study participants are randomly assigned to treatment and control groups within sites—give researchers a powerful method for understanding a program's effectiveness. Once they know the direction (favorable or unfavorable) and magnitude (small or large) of a program's impact, the next question is *why* the program produced its effect. Multi-site evaluations offer a chance to "get inside the black box" and explore that question.

First, researchers estimate the *overall* impact of the program without selection bias or other sources of bias, and then use cross-site analyses to connect program structure (*what* is offered) and implementation (*how* it is offered) to the magnitude of the impacts. However, these estimates are non-experimental and may be biased.

The CAMIC method takes advantage of randomization of a program component in only some sites to improve estimating the effects of other program components and implementation features that are not or cannot be randomized. The paper describes the method for potential use in the Health Profession Opportunity Grants (HPOG) program evaluation.

A simulation analysis of CAMIC shows that the method does not consistently reduce bias and, in some cases, increases bias. Nevertheless, we argue that presenting details of the method is useful. We urge other researchers to consider other settings where the method might be successfully applied in order to help evaluators learn more about what works.

Primary Research Question

Can the CAMIC method improve our ability to detect, without bias, which program components and implementation features are essential to a program's success?

Purpose

In job training evaluations, the program components (such as a given curriculum or support service) are rarely randomized to sites; and most implementation features (such as the dynamism of a site administrator) cannot be randomized. Instead, each site chooses its own configuration of program components to adopt and each possesses its own set of implementation features. As a result, the reasons that a particular combination of program components and implementation features exists in a site are also correlated with the program's impact. For example, the local program director's enthusiasm and leadership might be associated with both the choice of a particular program component and how well the component is implemented. Better implementation, in turn, might lead to greater program impact. But if that implementation feature is not measured and therefore is excluded from the researcher's cross-site attributional analysis, then estimates may overstate the influence of the program component on impact.

Multi-site experiments can facilitate an understanding of the effects of both program components and implementation features. This is the motivation for this work: to test whether a new method can help researchers better estimate the contributions of program components and implementation features to overall program impact.

Key Findings & Highlights

The paper describes how the CAMIC method uses the experimental estimate of the influence of a particular program component to specify the statistical model used for understanding the influence of other, non-randomized program components and implementation features. The goal is to identify the model that is least biased. However, the theoretical work demonstrates that there may be no single model that is the least biased for all estimates. Depending on the specific correlations among measured and unmeasured program components and implementation features, the model that produces the least biased measure of the influence of one program component may produce the most biased model of the influence of another.

Simulation work investigated how often the CAMIC method could select the least biased estimate of the influence of a particular program component. These simulations were unable to find generalizable conditions under which the CAMIC method is likely to reduce bias. Across all simulations, results were favorable to the CAMIC method for 47 percent of the parameters tested.

Methods

The Health Profession Opportunity Grants (HPOG) program's impact evaluation is assessing whether providing access to health sector career pathways training improves participant outcomes overall. To do this, individuals are randomized to the HPOG treatment group or to a control group that does not have access to HPOG-funded services. Importantly, in some sites there are *two* treatment groups, and this allows researchers to focus on a given program component's relative impacts. In particular, one treatment group has access to HPOG while the second treatment group has access to HPOG enhanced with one of three additional program components: facilitated peer support groups, emergency assistance for specific needs, or noncash incentives that encourage desirable program outputs and outcomes. These three studies can provide strong experimental evidence of the relative contribution of peer support, emergency assistance, or noncash incentives to the HPOG program's impact.

The CAMIC method is designed to exploit three-armed randomization of certain program components to estimate the effect of other components or intervention features through cross-site non-experimental attributional analysis. That is, can having experimental evidence on one of HPOG's experimentally evaluated enhancements improve our ability to gauge the effectiveness of other HPOG program components in situations where we observe these same program components naturally occurring in the program?

The method's basic approach involves the following:

1. Estimating the experimental impact of the program component (e.g., peer support) added to HPOG through a second treatment arm;

2. Calculating several alternative, non-experimental estimates of the impact contribution of the program component (e.g., peer support) from cross-site analysis models that control for various sets of site-level influences in each analysis;

3. Choosing the model that minimizes the difference between the experimental and non-experimental estimates of the impact contribution of the added program component (e.g., peer support); and

4. Applying that model to estimate the contribution to program impact of other program components (not peer support but other components naturally occurring in the HPOG program such as intensive case management or the presence of career pathways principles) or implementation features (e.g., the

program’s administrative structure, or case workers’ client orientation) when these other components or features do not vary randomly among individuals within sites or across sites.

CONTENTS

1. Introduction

Randomized experiments provide researchers with a powerful method for understanding a program's effectiveness. Once they know the direction (favorable or unfavorable) and magnitude (small or large) of a program's impact, the next question is *why* the program produced its effect. Programs that operate in many locations—and the multi-site evaluations that accompany them—offer an opportunity to "get inside the black box" and explore that question. That is, what is it about how the program is configured and implemented that leads its impact to vary?

Multi-site experiments—in which study participants are randomly assigned to treatment and control groups within sites—do this by enabling researchers first to estimate the *overall* impact of the program at each site without selection bias or other sources of bias, and then to move on to cross-site attributional analyses that connect the specifics of program configuration (*what* is offered) and implementation (*how* it is offered) to the magnitude of the impacts.

A small but growing portion of the literature evaluating the impacts of social programs uses multi-site experiments to investigate what it is about a particular program that determines its impacts (e.g., Bloom, Hill, & Riccio, 2001; 2003; Dorsett & Robins, 2013; Godfrey & Yoshikawa, 2012; Greenberg, Meyer, & Wiseman, 1994).

For example, in the Health Profession Opportunity Grants (HPOG) program impact evaluation, more than 10,000 individuals have been randomized to gain access to HPOG-funded services in almost 40 program locations. Looking across all these individuals and locations will provide an overall assessment of HPOG's impact; but examining how cross-site variation—in both the *what* and the *how*—associates with variation in program impacts can help inform lessons for future program design and practice. In HPOG, the *what* (which in this paper we call "program components") is health sector, career pathways-based education, training, and supports. The *how* (what we call "implementation features") pertains, for example, to the office culture at a site, the autonomy of its staff, or their client-centered orientation.

In a program evaluation, the program components (such as a given curriculum or support service) themselves are rarely randomized to sites; and implementation features (such as the dynamism of a site administrator) cannot be randomized. Instead, each site chooses its own configuration of program components to adopt and each possesses its own set of implementation features. For example, site planners taking up the HPOG program might decide to offer HPOG's facilitated peer support as part of their job training; and their own traits that associate with that decision (such as the administrator's leadership ability) will inform how the program is implemented in practice. Thus, non-experimental attribution of impact to various aspects of the site can be biased when the choices that sites make of what to offer and how to offer it reflect underlying factors that are hard to account for in the analysis—but that themselves may cause program impacts to be larger or smaller.

The reasons that a particular combination of program components and implementation features exists in a site also are correlated with the program's impact. For example, the local program director's enthusiasm and leadership might be associated with both the choice of a particular program component and how well the component is implemented. Better implementation, in turn, might lead to greater program impact. But if that implementation feature is not measured and therefore is excluded from the researcher's cross-site attributional analysis, then estimates may overstate the influence of the program component on impact. Likewise, the local environment—including economic conditions and policy context—might be associated with both the program components adopted by the site and the routes by which program participants can benefit from it, such as the number of "career ladder" jobs in the local economy. This type of contextual influence also may be difficult to measure and control for in the analysis. These and other such scenarios would bias estimates of the influence of program components and implementation features when they do not vary randomly across sites.

It is possible to randomize individuals to gain access to program components (for example, a lottery can be used in a job-training program to decide which participants are offered internships). But it generally is not possible to randomize to implementation features. That said, research is interested in understanding the effects of both program components and implementation features, and multi-site experiments can facilitate that learning.

1.1 Introducing the CAMIC Method

This paper examines whether having an experimental estimate of the contribution that a program component makes to impact magnitude can reduce the bias in estimating the contribution that other program components (or implementation features) make across sites that *choose* how to configure and implement their programs (rather than those components or features being assigned randomly to them).

In doing that, the paper considers studies with a strong foundation for attributing impacts to the program in general: randomization of individuals in each site to either program access (the treatment group) or total program exclusion (the control group). It investigates whether adding a second randomly assigned treatment arm that offers an enhanced version of the HPOG program (that includes a program component not available as part of the standard program offered to the first treatment arm) can improve non-experimental estimates of the impact contribution of other program components or of implementation features.

For example, in the case of the HPOG program, a three-armed randomized experimental evaluation is being implemented in some but not all sites, and for some but not all program components. That evaluation is assessing whether providing access to health sector career pathways training improves outcomes overall. To do this, individuals are randomized to the HPOG treatment group or to a control group that does not have access to HPOG-funded services. In some sites, there are two treatment groups. One group has access to HPOG, while the second treatment group has access to HPOG enhanced with one of three additional program components: facilitated peer support groups, emergency assistance for specific needs, and noncash incentives that encourage desirable program outputs and outcomes. These three experimental tests will provide strong evidence of the relative contribution of peer support, emergency assistance, or noncash incentives to the HPOG program's impact.

The new method that this paper describes is designed to exploit three-armed randomization of certain program components to estimate the effect of other components or intervention features through cross-site non-experimental attributional analysis. That is, can having experimental evidence on one of HPOG's experimentally evaluated enhancements improve our ability to gauge the effectiveness of other HPOG program components in situations where we observe these same program components naturally occurring in the program world? We refer to the method as CAMIC, for Cross-Site Attributional Model Improved by Calibration to Within-Site Individual Randomization Findings.

The CAMIC method seeks to reduce bias by identifying measures that account for site-level selection of program design and implementation to include in the statistical model. The number of measures an analyst can include in the model is limited by the number of sites in the analysis. That is, the analysis cannot include more measures than the number of sites; instead, the rule of thumb is that about one measure per eight sites is the most that should be included. This "degrees of freedom" limitation implies that analysts need to be selective in what enters into the analytic model; and the CAMIC method provides one way to be selective.

The method's basic approach involves the following:

5. Estimating the experimental impact of the program component (e.g., peer support) added to HPOG through a second treatment arm;

6. Calculating several alternative, non-experimental estimates of the impact contribution of the program component (e.g., peer support) from cross-site analysis models that control for various sets of site-level influences in each analysis;

7. Choosing the model that minimizes the difference between the experimental and non-experimental estimates of the impact contribution of the added program component (e.g., peer support); and

8. Applying that model to estimate the contribution to program impact of other program components (not peer support but other components naturally occurring in the HPOG program world such as intensive case management or the presence of career pathways principles) or implementation features (e.g., the program's administrative structure, or case workers' client orientation) when these other components or features do not vary randomly among individuals within sites or across sites.

1.2 About This Paper

This paper examines the CAMIC method as a potential innovation to the evaluator's toolkit: Can experimental evidence from a three-armed experiment be leveraged to reduce bias in non-experimentally estimated impact estimates? The intuition is strong: If we know the "truth" from the experimental evidence, then that should help in bringing non-experimental estimates closer to that truth. The CAMIC method uses the difference between experimental and non-experimental impact estimates to choose the model results in the least-biased non-experimental estimates, and then extends model to estimate the contribution to program impact of other program components or implementation features.

The method was developed for use in the HPOG Impact Study, where we will have both experimental and non-experimental evidence about the effectiveness of some of the program's components. While we wait for that study's data to become available, this paper undertakes a simulation study to examine the CAMIC method's properties.

In brief, despite the intuitive promise of the CAMIC method, in the simulations performed to date, the method does not consistently reduce bias in estimates of the contribution to program impact of components or features that were not randomized. We believe that presenting details of the method are useful regardless, so that future simulations and future applied analytic tests can continue to consider the approach and contribute to advancing program evaluation methodology.

The paper proceeds as follows. The next section describes, both conceptually and analytically, the standard approach to producing cross-site estimates of the impact contributions of program components and implementation features, when individuals are randomly assigned to treatment and control groups within sites. Then we extend this framework to the situation in which we have an additional experimental treatment arm capable of isolating the effect of a particular program component or implementation feature. We explain how the CAMIC method can use data from this kind of three-armed experiment to estimate the impact of *all* the components and features of interest. A simulation exercise examines the circumstances in which the CAMIC method leads to less-biased non-experimental impact estimates. Finally, we summarize the paper's contributions and make suggestions for future research in this area.

2. Standard Approach to Cross-Site Attributional Analysis with Experimental Data

This section explains that various site-level factors—which may be measured or unmeasured—can influence the magnitude of a program's impact, and it provides an analytic model for deriving these "moderating" influences from experimental data on many sites whose approaches to the program vary.

It begins by identifying the types of moderators often considered as potential causes of variation in impact magnitude across the sites of a multi-site social program experiment. It then describes the multi-level analytic framework for attributing variation in impact to program components, implementation features, and other site-level explanatory factors. This multi-level model is built on the structure of the HPOG Impact study, where individuals are randomly assigned to treatment and control groups within sites, but sites are free to choose how to configure and operate their programs rather than having program components or implementation features randomly assigned to them.

Consider HPOG again: There are a lot of similarities and a lot of differences among the many HPOG sites, and this variation may lead to variation in the program's overall impact. The evaluation's first research question considers the average impact across all the sites and all types of individuals. What the study is also interested in understanding, however, is the extent to which certain program components are *essential* ingredients to the HPOG recipe. As noted, the HPOG study has some three-armed experimental sites, which will examine the relative contribution of particular program components to HPOG's impact. The study will be able to do so both through the experimental evidence alone (in the three-armed programs) and through the cross-site variation that exists naturally in the HPOG program world.

In analyzing the impact of these selected program components where access to HPOG is determined through a lottery (randomly), the analytic model provides an experimental estimate of the intervention's impact in each site that is statistically consistent.[1] In analyzing the impact of selected program components and implementation features where access to HPOG is not random, the analytic model estimates the influence of these components and features on the magnitude of the program's overall impact when the components and features vary from site to site.[2]

2.1 Types of Moderators

A key reference in this field is Bloom, Hill, and Riccio (2003), which hypothesizes site- and individual-level factors that could affect the magnitude of intervention impacts. As detailed in the text box that follows, these include four types of site-level factors—*program components, implementation features,*

[1] "Statistically consistent" means that the impact estimate moves extremely close to the true impact in the population as the sample size gets very large.

[2] Unlike the lottery estimate, these latter estimates may be subject to omitted-variable bias (and therefore may be inconsistent in very large samples) because the program components and implementation features of interest are not randomly assigned to sites. Instead, the sites *choose* what to implement and how to implement it. Additional explanation regarding why this is a problem appears in Appendix A.

local context, and *participant composition*—which may be measured or unmeasured. The factors underpin our framework for thinking about HPOG's impacts.

Site-Level Factors that May Influence Program Impacts

- *Program Components.* Program components (or program "activities," as Bloom, Hill, and Riccio refer to them) represent the services offered to program participants. For example, in a job-training program, they include activities such as job search assistance and vocational training.
- *Implementation Features.* Implementation features describe the practices and views of administrators and staff operating the program. For example, Bloom, Hill, and Riccio used the following variables when analyzing the implementation of welfare-to-work programs: the degree of a program's emphasis on moving clients into jobs quickly, the degree of personalized attention given to clients, the closeness of client monitoring, frontline staff and staff/supervisor inconsistency in views about the agency's service approaches, and staff caseload size.
- *Local Context.* Site-level local context represents the environment in which the site is located. Relevant factors might include characteristics related to the economy (e.g., the unemployment rate), crime, housing market, demographic characteristics, or other relevant measures of the social, political, and economic climate.
- *Participant Composition.* Impact magnitude might vary for various types of clients or be influenced by the composition of the clients being served. The aggregation of characteristics might include participants' demographic, education, and economic backgrounds, as well as household traits (e.g., marriage status) and composition (e.g., number of children).

Multi-level analyses of experimental data tend to control for local context and participant composition so that analyses can focus on the impact of selected program components and implementation features. As described in further detail below, the goal is to estimate the contribution of each of the selected program components or implementation features to the magnitude of the program's overall impact.

2.2 Analytic Model

When individuals are clustered within sites in a multi-site evaluation, it is customary to use a multi-level model to estimate the relationship between program impacts and the relevant site-level measures, as described above. The following two-level model depicts program impacts as a function of individual-level and site-level measures.

The unit of analysis for Level One is the individual member of the study sample, while the unit of analysis for Level Two is the site.

Level One: Individuals

$$Y_{ji} = \alpha_j + \beta_j T_{ji} + \sum_c \delta_c IC_{cji} + \sum_c \gamma_c IC_{cji} T_{ji} + \varepsilon_{ji} \qquad \text{(Eq. 1)}$$

Level Two: Sites

$$\beta_j = \beta_0 + \sum_m \pi_m P_{mj} + \sum_g \varphi_g I_{gj} + \sum_d \tau_d PC_{dj} + \sum_q \zeta_q LC_{qj} + \mu_j \qquad \text{(Eq. 2)}$$

and

$$\alpha_j = \alpha_0 + \sum_q \kappa_q LC_{qj} + v_j \qquad \text{(Eq. 3)}$$

Combining the elements of the above two-level model produces the following:

$$Y_{ji} = \alpha_0 + \sum_q \kappa_q LC_{qj} + \beta_0 T_{ji} + \sum_m \pi_m P_{mj} T_{ji} + \sum_g \varphi_g I_{gj} T_{ji} + \sum_d \tau_d PC_{dj} T_{ji} + \sum_q \zeta_q LC_{qj} T_{ji} + \sum_c \delta_c IC_{cji} + \sum_c \gamma_c IC_{cji} T_{ji} + \{v_j + \mu_j T_{ji} + \varepsilon_{ji}\} \qquad \text{(Eq. 4)}$$

In these equations, Y is the outcome of interest, indexed by i individuals and j sites. In Equation (4), program components (P_{mj}), implementation features (I_{gj}), participant composition measures (PC_{dj}), and local context measures (LC_{qj}) are all multiplied by the treatment indicator. These interaction terms capture the influence of a given site-level measure on the magnitude of the program's impact. Local context measures enter the model directly to capture the influence of the environment on the outcomes of those individuals in the treatment and control groups. This specification includes individual-level characteristics (IC_{cji}) that affect these outcomes for both groups; it also includes participant composition, program components, and implementation features, which affect these outcomes only for individuals in the treatment group (because control group members never come in contact with the program or with other program participants). See Exhibit 1 for definitions of the terms included in these equations and all models presented throughout the manuscript.

Ultimately, the goal is to discover the causal impact on participant outcomes of adding a program component or implementation feature not already incorporated in participants' experience of the program. For example, in the HPOG Impact Study, we are analyzing the impact of offering program components such as emergency assistance and intensive case management. We are also analyzing the impact of implementation features, such as the extent to which the program operates on the principles of the career pathways framework or the extent to which program staff emphasize education or employment. The evaluation's Analysis Plan discusses additional detail regarding the specific measures of interest that can be analyzed in this framework (see Harvill et al., 2015).

In the equations above, estimates of the π_m coefficients—call them $\hat{\pi}_m^N$ (typically estimated using maximum likelihood methods, such as those described in Bryk and Raudenbush (1992))—are intended to capture the causal connection between program components and impact magnitude. The N superscript denotes that this estimate is non-experimental, because it is computed using cross-site variation in program components. Similarly, the estimates $\hat{\varphi}_g^N$ of the coefficients φ_g are intended to capture the causal connection between implementation features and impact magnitude.

Limited degrees of freedom for this analysis, which are determined by the number of Level Two units, constrain the number of Level Two measures we can include in the model. As a result, the process of choosing which measures to include becomes quite important.

Because estimates of these coefficients are identified by the natural variation in program components and implementation features across sites, omitted-variable bias arises in this analysis if one or more site-level measures exist that influence impact magnitude; that are not included in the Equation (4) model; and that correlate with one or more of the site-level measures that are included. This makes estimates of all the site-level coefficients, including $\hat{\pi}_m^N$ and $\hat{\varphi}_g^N$, non-experimental and opens up the possibility of policy conclusions confounded by extraneous influences. Appendix A includes a detailed description of circumstances that can cause omitted-variable bias in estimates of the contributions of program components and implementation features to site-level impacts.

Exhibit 1: Definition of Model Terms

Term	Definition
Outcome and Covariates	
Y_{ji}	The outcome measure for individual *i* from site *j*
T_{ji}	The standard treatment group indicator (1 for those individuals assigned to the standard treatment group, and 0 for those individuals assigned to the enhanced treatment group or the control group; this is labelled "T" for "treatment")
E_{ji}	The enhanced treatment group indicator (1 for those individuals assigned to the enhanced treatment group, and 0 otherwise; this is labelled "E" for "enhanced" treatment)
TE_{ji}	The treatment group indicator (1 for those individuals assigned to the standard treatment or enhanced treatment group, and 0 for those individuals assigned to the control group; this is labelled "TE" for the combination of standard "treatment" and "enhanced" treatment)
IC_{cji}	Individual baseline characteristic *c* for individual *i* from site *j*, *c* = 1, . . ., *C* (these are labelled "IC" for "individual characteristics")
P_{mj}	Program component *m* for site *j*, *m* = 1, . . ., *M* (these are labelled "P" for "program")
I_{gj}	Implementation feature *g* for site *j*, *g* = 1, . . ., *G* (these are labelled "I" for "implementation")
PC_{dj}	Participant composition variable *d* for site *j*, *d* = 1, . . ., *D*; this is a site-level aggregation of the individual characteristics (ICs) (these are labelled "PC" for "participant composition")
LC_{qj}	Local context variable *q* for site *j*, *q* = 1, . . ., Q (these are labelled "LC" for "local context")
Model Coefficients	
α_j (alpha)	The control group mean outcome (counterfactual) in site *j*
β_j (beta)	The conditional impact of being offered the standard intervention for each site *j*
δ_c (delta)	The effect of individual characteristic *c* on the mean outcome, *c* = 1, . . ., *C*
γ_c (gamma)	The influence of individual characteristic *c* on impact magnitude, *c* = 1, . . ., *C*
β_0	The grand mean impact of the standard treatment
π_m (pi)	The influence of program component *m* on impact magnitude, *m* = 1, . . ., *M*
φ_g (phi)	The influence of implementation feature *g* on impact magnitude, *g* = 1, . . ., *G*
τ_d (tau)	The influence of participant composition variable *d* on impact magnitude, *d* = 1, . . ., *D*
ζ_q (zeta)	The influence of local context variable *q* on impact magnitude, *q* = 1, . . ., Q
α_0	The grand mean control group outcome

Term	Definition
κ_q (kappa)	The effect of local context variable *q* on control group mean outcome, *q* = 1, . . ., *Q*
π_{ej}	The impact of being offered an enhanced program that includes component *e* relative to the standard program for each site; this and the other subscripted πs are program component impacts
π_e	The grand mean impact of being offered the enhanced intervention inclusive of component *e*, rather than the standard intervention without *e*
Error Terms	
ε_{ji} (epsilon)	A random component of the outcome for each individual
μ_j (mu)	A random component of the standard intervention impact for each site
ν_j (nu)	A random component of the mean outcome for each site
ω_j (omega)	A random component of the enhanced intervention's incremental impact for each site

2.3 Addition of Three-Armed Sites

Consider next the situation where some sites randomize individuals into not just treatment and control groups but also into a third "enhanced" treatment group. Those assigned to this third arm are offered the standard treatment plus an additional program component—the "enhancement." This design allows us to generate an experimental estimate of the enhancement's impact. Exhibit 2 portrays two alternative examples of how a six-site experiment might allocate selected program components to sites.

Left-side example (two-armed). The example on the left side of Exhibit 2 depicts an experiment with two-armed randomization of individuals between a control group and a treatment group. Each site has designed its own version of a "standard" program, choosing which components to include and how to implement them. Program offerings vary naturally from site to site: Two of the sites (F and G) offer their version of a standard program (which may differ between sites in ways not shown); two of the sites (H and I) offer their version of a standard program that contains component #1; and two of the sites (J and K) offer their version of a standard program that contains component #2.

Right-side example (three-armed). The right side of Exhibit 2 depicts an alternative configuration of the same number of sites. Again, two of these sites (L and M) offer their own version of a "standard" program. Another two sites offer their version of a standard program, where the program at site N contains component #1 and the program at site O contains component #2. What is new is that two sites have added a second treatment configuration: One site (P) randomizes individuals to either its standard program (the T_S arm) or its standard program enhanced by the addition of component #1 (the T_E arm); and one site (Q) randomizes individuals to either its standard program or its standard program enhanced by the addition of component #2.

The right side of the exhibit represents a simplified version of the HPOG program and its evaluation where—across 42 sites—several sites randomize individuals among a control group (excluded from the program), a "standard" program (first treatment arm), and a second treatment arm that includes the standard program plus one of three enhancement components. In HPOG, the three enhancements are peer support, emergency assistance, and noncash incentives.

Note that these program enhancements also exist naturally in two-armed sites—in this simplified diagram and in the actual HPOG evaluation.

Exhibit 2. Illustrative Six-Site Experimental Designs

Six Two-armed Sites		Four Two-armed and Two Three-armed Sites	
Control/Treatment	**Added Program Component**	**Control/Treatment**	**Added Program Component**
F. C T		L. C T	
G. C T		M. C T	
H. C T	Naturally Occurring Component #1	N. C T	Naturally Occurring Component #1
I. C T	Naturally Occurring Component #1	O. C T	Naturally Occurring Component #2
J. C T	Naturally Occurring Component #2	P. C T_S T_E	Randomized to Component#1
K. C T	Naturally Occurring Component #2	Q. C T_S T_E	Randomized to Component #2

Under the standard multi-site, multi-level analysis of the two-armed sites represented on the left side of Exhibit 2, the analysis would compare the treatment-control differences in mean outcomes in sites H and I with the treatment-control differences in mean outcomes of individuals in all the other sites (accounting for site-level contextual variables). This comparison would determine the contribution of program component #1 to overall program impact. This is a non-experimental analysis in which the experimental treatment-control difference (impact) in sites H and I may differ from differences in the other sites for reasons other than the presence of program component #1.

In practice, this analysis would be carried out in a multiple regression framework as detailed in Section 2.2. Here we discuss the analysis conceptually, however, to make clear the between-site comparisons being made to estimate impacts. In the “flat” regression provided by Equation (4) above, if other included site-level moderators of impact are sufficient and convincing at eliminating confounding factors, then the interpretation of the estimated impact of component #1 as causal and unbiased is more likely, but will never be complete.

The same type of analysis can take place with the two-armed sites represented on the right side of Exhibit 2, where sites N and O inform the non-experimental analyses. In addition, sites P and Q can be used for generating an experimental estimate of the effect of program components #1 and #2, respectively, through a comparison between their two treatment arms. As such, they provide an opportunity to “calibrate” the non-experimental estimates of those same two program components. Because we have the “right” answer to the questions of how these two components contribute to program impacts from these experimental comparisons, we can vary the other site-level moderators included in the model in Equation (4) until these non-experimental estimates are as close as we can get them to their corresponding experimental estimates.

3. The CAMIC Method

This section describes the CAMIC method for selecting the set of site-level covariates to include as impact moderators when seeking to attribute cross-site impact differences to the program components and implementation features adopted by local social service agencies. We begin by describing the intuition motivating the CAMIC method. Then, we specify how to estimate the experimental impact of the program component (e.g., peer support) added to HPOG though a second treatment arm. We then describe the steps in the CAMIC approach that calibrate the cross-site attributional model described in Equations (1) through (4) in the previous section: estimate these equations using a variety of models that include different sets of site-level measures; select the model that minimizes the difference between the experimental and non-experimental estimates of the impact contribution of the added program component (e.g., peer support); and apply that model to estimate the contribution to program impact of other program components (not peer support but other components naturally occurring in the HPOG program world such as intensive case management or the presence of career pathways principles) or implementation features (e.g., the program's administrative structure, or case workers' client orientation).

Throughout the discussion, we assume that the model always includes a set of priority program components and implementation features. These might be prioritized because they interest policymakers, practitioners, or researchers.

Beyond these priority program components and implementation features, the goal is to select the best set of measures (those contextual factors discussed earlier) to analyze in order to reduce the degree to which other influences on impact magnitude confound estimated effects of the program components and implementation features that are the focus of the analysis. The context measures for this purpose include local community characteristics, the collective characteristics of program participants at a site (to control for possible peer effects on impact), and further program components and implementation features not already included in the model (i.e., not of primary interest).

3.1 Building Intuition for the CAMIC Method

The CAMIC method identifies the non-experimental model that most closely reproduces the experimental result for a particular program component. We hypothesize that this model will also produce less-biased estimates of the impact contributions of other non-randomized program components and/or implementation features. Whether this proves true will depend on how much commonality there is among the omitted factors that influence the choices sites make in *what* program components to offer and *how* to implement them.

Bias stems from unobserved factors that influence both the choices that define the program and the program impact directly. If the same unobserved factor—such as the talent of the program administrator—affects the program's choice to offer (or not offer) program components #1 and #2, then the bias in the estimates of the impact contributions for these two components is related. In this case, reducing the bias in the estimate of the impact contribution of program component #1 may reduce the bias in the estimate of the impact contribution of program component #2.

Alternately, there may be multiple unobserved factors that are related to program effectiveness but unrelated to one another; these include, for example, the skill of the case managers and the quality of the instruction offered. If the skill of the case managers is related to program component #1 and the quality of instruction is related to program component #2, then the bias in the two program components is coming from different sources. Reducing the bias in program component #1 involves controlling for the skill of the case managers, which does not reduce the bias in program component #2 which stems from unobserved quality of instruction.

Most likely, both scenarios are true to some extent. We expect that there is an unobserved factor that affects bias in estimates of impact contributions for all program components and implementation features, and there are also additional unobserved factors that are related to only a few of the measured components and features. How much reducing the bias in the estimate of the impact contribution of program component #1 can reduce the bias in the impact contribution of program component #2 could be investigated using HPOG data. We will be able to calculate an experimental estimate of the impact contribution of both facilitated peer support and emergency assistance. We could apply the CAMIC method to minimize the bias in the contribution of facilitated peer support and then compare the non-experimental estimate of the impact contribution of emergency assistance to the experimental estimate. Until the HPOG data are available to the study team, this paper provides a test by use of simulations.

3.2 Deriving the Experimental Benchmark

In this section, we describe how researchers can use three-armed random assignment to experimentally estimate the impact of a program component offered as an enhancement in the sites with the additional, third treatment arm. Notationally, consider the experimental design summarized in Exhibit 2 (right side), where individuals in a subset of sites j =1, ... , J^* ($J^* < J$) are randomly assigned to one of three arms: a standard treatment group, an enhanced treatment group (that receives the standard treatment plus an enhancement component), or a control group (that has no access to the program). In all other sites j = J^*+1, ... , J, individuals are randomized to just two arms: a standard treatment group (where the treatment does not include the enhancement) and a control group.

The experimental estimate of the impact of the enhancement can be computed under these circumstances using a two-level model and an analysis sample limited to sites j =1, ... , J^* with three-armed random assignment. The Level One regression equation depicted by Equation (5) below—which parallels the earlier Equation (1) with modifications—uses data on individuals in site j to model the relationship between an outcome Y and an overall treatment indicator (which denotes whether the participant was assigned to either standard treatment or enhanced treatment) and an enhanced treatment indicator, while controlling for individual characteristics. The impact coefficients of interest in this equation (β_j and π_{ej}) and the control group mean (α_j) in each site serve as the dependent variables for Level Two of the model, as depicted in Equations (6), (7), and (8). Exhibit 1 defines these terms.

Level One: Individuals

$$Y_{ji} = \alpha_j + \beta_j TE_{ji} + \pi_{ej} E_{ji} + \varepsilon_{ji} \qquad \text{(Eq. 5)}$$

Level Two: Sites

$$\beta_j \ = \beta_0 + \mu_j \qquad \text{(Eq. 6)}$$

$$\pi_{ej} \ = \pi_e + \omega_j \qquad \text{(Eq. 7)}$$

and:

$$\alpha_j = \alpha_0 + \ v_j \qquad \text{(Eq. 8)}$$

We can simplify the above two-level model by substituting Equations (6), (7), and (8) into Equation (5), which produces the following single equation:

$$Y_{ji} = \alpha_0 + \beta_0 TE_{ji} + \pi_e E_{ji} + \{v_j + \mu_j TE_{ji} + \omega_j E_{ji} + \varepsilon_{ji}\} \qquad \text{(Eq. 9)}$$

Estimating Equation (9) through linear regression, we obtain—among other things—an estimate $\hat{\pi}_e^X$ of π_e straight from the experiment, based on purely random variation in which individuals receive a program that includes the enhancement element *e* and which individuals do not. The *X* superscript denotes the unbiased experimental nature of this estimate.

3.3 Using the Experimental Estimate to Calibrate the Non-Experimental Model

The CAMIC method selects the set of impact moderators that produces the smallest measured difference between the experimental estimate of the impact contribution of the enhancement, $\hat{\pi}_e^X$ above, and a non-experimental estimate of the impact contribution of that same component.

The first step in implementing this method has just been described: experimentally estimating the effect of the enhancement in the three-armed sites. From there, applying the CAMIC approach produces non-experimental estimates of the impact contribution of that program component and other site-level moderators by including different combinations of site-level measures as covariates in the analytic model. In its final steps, the CAMIC method generates the set of site-level covariates that minimizes the measured difference between the experimental and non-experimental estimates of the impact of the enhancement—and uses the *same* site-level covariates to produce non-experimental estimates of the contribution of *other* program components or implementation features to the program's overall impact.

Those steps proceed as follows:

Step 1. Compute an experimental estimate of the impact of the enhancement ($\hat{\pi}_e^X$) using data from sites that conduct three-armed random assignment.

Step 2. Compute a non-experimental estimate of the impact of the enhancement ($\hat{\pi}_e^N$) by estimating Equation (4). To produce this estimate, the sample is limited to (1) the control and enhanced treatment arms from sites that conducted three-armed random assignment using the enhancement and (2) the control and standard treatment arms from sites that did not use the enhancement. Sample members randomly assigned to the standard treatment arm in sites that conduct three-armed random assignment are excluded from this analysis to "pretend" that these sites had chosen to use that same component *e* as part of their standard program.[3] This forces us to estimate the effect of program component *e* as we would without randomization of its use between two experimental arms. The Level Two moderators in Equation (4) help remove confounding bias from the $\hat{\pi}_e^N$ estimator.

Step 3. Analyze all combinations of variables (subject to degrees of freedom limitations) in Equation (4) to find the set of Level Two moderators that produces the non-experimental estimate of the enhancement's contribution to program impact with the least measured bias, where bias is measured subject to sampling variability as $|\hat{\pi}_e^N - \hat{\pi}_e^X|$. Potential variables for this bias reduction exercise include program components and implementation features of secondary interest, participant composition measures, and local context measures.[4]

Step 4. Compute cross-site estimates of the contributions of program components to impact magnitude ($\hat{\pi}_1^N$,..., $\hat{\pi}_e^N$,..., $\hat{\pi}_M^N$) and the contributions of implementation features to impact magnitude ($\hat{\varphi}_1^N, \dots, \hat{\varphi}_G^N$) while controlling for the set of moderators selected in Step 3, using maximum likelihood estimation methods from Bryk and Raudenbush (1992). This estimation uses the entire sample, across all sites and including individuals from all three randomization arms to gain greater statistical precision in all the coefficient estimates, including the estimates of the components' and features' contributions to impact magnitude.

To incorporate the full sample, modifications are made to the earlier cross-site attributional model that resulted in Equation (4) earlier. These modifications are detailed in Appendix B and add terms to Equation (4) to create in a full-sample analysis that includes both two-armed and three-armed sites a distinction between the impact contributions of program components in the standard treatment arms of the various sites and the impact contribution of the enhancement in the three-armed sites.

[3] This necessitates replacing T_{ji} in Equation (4)—which distinguishes between standard treatment group members and control group members—with TE_{ji} (see Exhibit 1 in Section 2 for definitions), which distinguishes between enhanced treatment group members of all sorts and all control group members. P_{ej} also needs to be replaced with $P*_{ej}$, which equals P_{ej} in the two-armed sites that are used at this step (which, by construction, all have $P_{ej} = 0$) and equals 1 in the three-armed sites.

[4] Note that adding a covariate as an impact moderator to an attributional model may actually increase omitted variable bias, even if the added variable is highly correlated with omitted confounders (Steiner & Kim, 2015). One goal of the CAMIC method is to avoid this mistake. This perverse result can arise from two phenomena: (1) bias amplification and (2) removing the benefit of offsetting biases. Bias amplification occurs when conditioning on the new variable amplifies the bias caused by the omitted, unobserved confounder by increasing the correlation between the unobserved confounder and other included variables of interest. It is also possible that two omitted confounders initially induced bias in opposite directions, and that the benefit of these offsetting biases is lost when one but not both confounders is added to the specification.

4. Simulation Exercise

The CAMIC method is used to leverage experimental evidence to reduce bias in non-experimental estimates of the influence of program components and implementation features on a program's total impact. To investigate whether the CAMIC method might help accomplish this goal, we conduct simulations that explore the method in a simplified theoretical framework.

The simplification is possible because the relationships of interest here are at the site level. Random assignment of individuals within sites is important in a multi-site trial because it allows us to calculate site-level impacts. However, once that is done, we can explore the relationship between the variation across sites in measured impact and the variation across sites in a range of measures that may influence impact. Those measures include the program components and implementation features used in the program in any site, local context measures, and indicators of the composition of program participants. For example, such an exploration can be undertaken by estimating a site-level regression that expresses estimated impact in a site as a linear function of site-level measures of all these measures.

A more sophisticated multi-level modeling approach, such as that described above in Equations (1) through (3), integrates these steps and provides correct standard errors for hypothesis testing. As a result, it is the preferred approach to analyzing data in practice. However, to explore the CAMIC method, we can focus on just the site-level regression. This captures the relationships of interest between impacts and the impact moderators in the various categories noted. It also captures the source of bias in the standard non-experimental measures of the effects of those moderators: omitted causal factors at the site level.

Bias arises when program components and implementation features are correlated with site-level factors that also influence impacts but that are omitted from the analysis. These factors—often omitted because they are unobserved in the data—may include aspects of the program that administrators choose (e.g., an unmeasured program component or implementation feature), aspects of the program context that are beyond their control (e.g., the local unemployment rate), and other unobserved factors that influence both the program components and implementation features and the impacts of the program (e.g., the raw talent of the program leadership). For all of these types of unobserved factors, if program components and implementation features are correlated with unobserved factors, the estimated influence of those components and features will reflect the influence of the unobserved factors.

The simplified framework used for our simulations focuses on the site-level relationship between the true impact of a program (Δ) and three program components (P_1, P_2, P_3), all of which are correlated with an unobserved, site-level factor (μ).[5] This relationship is given by

$$\Delta_j = \pi_0 + \pi_1 P_{1j} + \pi_2 P_{2j} + \pi_3 P_{3j} + \mu_j + \varepsilon_j, \qquad \text{(Eq. 5)}$$

where:

[5] Although we refer to the observed, site-level factors as program components here, they could be recast as any combination of program components, implementation features, local context, and participant composition without changing the approach or conclusions.

- j indexes programs; $j = 1,2, \ldots ,J$,
- Δ_j is the impact of the program implemented by site j,
- P_{mj} is a continuous measure of the extent to which program j implemented program component m,
- π_m is the influence of program component m on the program's impact,
- π_0 is the mean impact of the program,
- μ_j is a site-specific, unobserved factor that is correlated with observed program components, and
- ε_j is an error term unrelated to the observed program components.

In Appendix C, we derive an expression for the bias in the estimates of $\pi_0, \pi_1, \pi_2, \pi_3$. The estimate of the constant is unbiased. The bias in the coefficients of program components is the statistical expectation of a non-linear function of (1) the observed variance for each program component, (2) the observed covariance between each of the program components, and (3) the realized (but unobserved) covariance between each program component and the unobserved factor. The bias does not depend on the true influence of the program components on impact or on the variance of the unobserved factor.

In the expanded expression of bias in Appendix C, we see that bias is ultimately due to the correlation between program components and the omitted factor: each term in the numerator includes the covariance of one of the program components and the omitted factor. We can think of the correlation between the first program component and the omitted factor as the direct source of bias in estimates of the influence of that program component. If the covariance between the program components is set to 0, then the expression for bias simplifies to include only this direct effect. However, when the program components are correlated with one another, the bias in the first program component is also affected by the correlation between the omitted factor and the other two components. This is because the bias in the first program component is indirectly affected by the omitted factor through its correlation with the other program components.

When all program components are correlated with one another and with the omitted factor—as is most likely the case in any real-world application—the expression for bias does not yield simple statements about when bias will be larger or smaller. This is because the effect of increasing the value of a particular correlation depends on the values of all the other terms. For example, increasing the correlation between program component #1 and the omitted factor would increase the magnitude of the terms in which it appears. However, when those terms are added to the other terms, it might reduce bias in program component #1 if, say, the remaining terms were of opposite sign and the two sources of bias offset each other.

4.1 The CAMIC Method in a Simplified Simulation Framework

Suppose that we observe an unbiased measure of the true value of π_1 and that our goal is to obtain an estimate of π_2. In this case, the CAMIC method uses this unbiased estimate of π_1 to select the least biased specification between the following models:

Model 1: $\Delta_j = \pi_0^1 + \pi_1^1 P_{1j} + \pi_2^1 P_{2j} + \pi_3^1 P_{3j} + u_j^1$

Model 2: $\Delta_j = \pi_0^2 + \pi_1^2 P_{1j} + \pi_2^2 P_{2j} + u_j^2$

Note that we do not consider models that omit P_1 or those that omit P_2. Execution of the CAMIC method requires that P_1, the program component that was randomly assigned to sites, be included in the analysis. Furthermore, the model must include P_2 because the goal of the exercise is to obtain the least biased estimate of π_2, the coefficient on P_2. We refer to P_1 as the "reference component" and P_2 as the "focal component."

The error terms for these models include multiple terms. In the first model, the error term includes the unobserved factor:

$$u_j^1 = \mu_j + \varepsilon_j$$

The error term in the second model includes both the unobserved factor and the influence of the omitted program component:

$$u_j^2 = \pi_3 P_{3j} + \mu_j + \varepsilon_j$$

In Model 1, bias arises from the correlation between the omitted factor and the program components as discussed above. In addition, in Model 2, the omitted program component in the error term may contribute positively or negatively to the omitted-variable bias, yielding bias that is either larger or smaller than the bias from Model 1.

In Appendix C, we derive an expression for the bias in Model 2. This expression is similar to the one derived for Model 1—it is the statistical expectation of a non-linear function of many variables. However, the coefficient of the third program component affects bias in Model 2 because it appears in the error term and thereby affects the omitted-variable bias.

4.2 Simulation-Based Exploration of Bias

To understand the CAMIC method's potential to reduce bias in our estimate of π_2, we seek to understand whether the least biased specification for the randomized enhancement (P_1 in the above example) reference component is also the least biased specification for the target non-randomized program component (P_2 in the above example). Because we cannot take the expectation of the expression of the bias directly, we use simulation-based Monte Carlo integration to calculate the bias for particular sets of parameters to explore bias for a range of parameters.

Using Monte Carlo integration involves repeatedly generating values for random variables and calculating the value of the function for that value. After a large number of repetitions, the average of the observed value of the function gives the statistical expectation. Applying this process to calculating bias, we draw observations of the three program components and the omitted factor and calculate the bias in each model for the simulated dataset. Then, we calculate the mean bias in Model 1 and the mean bias in Model 2 across all the simulations. Finally, we consider whether the model that produces the least biased estimate of the reference component (π_1) also minimizes the bias in the focal component (π_2). We consider results favorable to the CAMIC method if the least biased model for π_1 is also the least biased model for π_2 and unfavorable otherwise.

We assume that the program components and the omitted factor are normally distributed with mean 0 and a standard deviation of 1. The key parameters that determine the bias are the correlations among the observed program components ($\rho_{12}, \rho_{13}, \rho_{23}$), the correlation of each observed program component with the observed factor ($\rho_{1\mu}, \rho_{2\mu}, \rho_{3\mu}$), and the true influence of the third program component on impact (π_3). To calculate bias, we must select a value for each of these seven parameter values.

Given the large number of possible combinations of parameter values, we must be strategic in selecting a relatively limited number of simulations that help us understand the range of possible biases. We first identify possible values for the observed parameters and refer to each of these values as a scenario. Then, for each scenario, we run 500 different simulations to capture a broad range of values of the unobserved parameters. We calculate the proportion of these simulations that are favorable to the CAMIC method for a particular scenario. This structure allows someone to consider which of our scenarios is most similar to the correlations they observe in their data.

Exhibit 3 describes the scenarios we investigate. We define sets of scenarios to answer the questions:

- How do the signs of the observed correlations affect the CAMIC method's potential?
- How does the overall magnitude of the observed correlations affect the CAMIC method's potential?
- How does the relative magnitude of the observed correlations affect the CAMIC method's potential?

Exhibit D.1 in Appendix D lists the full details for each of the 35 scenarios.

Exhibit 3. Characteristics of Scenarios Examined to Analyze the CAMIC Method's Potential

Scenarios	Focus of Exploration	Description
1-8	Sign	These scenarios set the magnitude of all correlations to 0.25 and systematically explore all possible combinations of sign for the three correlations.
9-21	Magnitude	These scenarios systematically increase the magnitude of the correlations while holding these correlations equal to one another. The scenarios move from ($\rho_{12} = \rho_{13} = \rho_{23} = 0.10$) to ($\rho_{12} = \rho_{13} = \rho_{23} = 0.70$).
22-29	Relative Magnitude	These scenarios change the correlation of one program component at a time and systematically explore all possible combinations of 0.25 and 0.50 as the value for the three correlations.
30-35	Relative Magnitude	These scenarios systematically explore correlations defined by all possible orderings of (0.25,0.50,0.70).

4.3 Simulation Findings

Exhibits 4 through 6 below present the results separately for three scenarios. The specific scenarios were selected to show the range of findings. Exhibit 4 displays results for the scenario with the least favorable findings for the CAMIC method; Exhibit 5 shows the scenario with the most favorable findings for the CAMIC method; and Exhibit 6 shows the scenario with the most typical findings in terms of favorability for the CAMIC method. Each exhibit comprises 20 color-coded panels in a grid of five rows and four columns. Altogether, we estimate the bias for 500 distinct specifications of unobserved parameters for each overarching scenario.

When presenting our findings for each scenario, we use the following color scheme:

🟩 least biased model for π_1 is also the least biased model for π_2 (green)

🟥 least biased model for π_1 is not the least biased model for π_2 (red)

The exhibits that follow show patterns of specifications that are favorable to the CAMIC method (green) and that are not favorable to the CAMIC method (red).

In each exhibit, the top left panel displays the results for the specifications with a particular value of the coefficient of P_3 and of the correlation between the reference component P_1 and the omitted factor μ: $\pi_3 = -0.50$ and $\rho_{1\mu} = -0.50$. The panel includes 25 different specifications that capture variation in the correlation between the focal component P_2 and the omitted factor μ, and in the correlation between P_3 and the omitted factor μ. The top left cell in the panel sets both these correlations $\rho_{2\mu} = \rho_{3\mu} = -0.50$. Moving right across the panel, the correlation between P_3 and the omitted factor μ increases to $\rho_{3\mu} = 0.50$. Moving down the panel, the correlation between P_2 and the omitted factor μ increases to $\rho_{2\mu} = 0.50$. In Exhibit 4, results are favorable to the CAMIC method (green) when the correlation between the focal component P_2 and the omitted factor μ is −0.50 and also when the correlation between focal component P_2 and the omitted factor μ is −0.25 and the correlation between P_3 and the omitted factor μ is 0.00 or greater.

Comparing columns of panels with one another isolates changes in the coefficient of $P_3(\pi_3)$. These changes affect the omitted-variable bias in Model 2 and have no effect on the bias in Model 1. Comparing rows of panels with one another isolates changes in the correlation between the reference component P_1 and the omitted factor μ ($\rho_{1\mu}$). Moving from the top row of panels to the middle row of panels, this correlation increases from $\rho_{1\mu} = -0.50$ to $\rho_{1\mu} = 0.00$.

For the middle row of panels, there is no direct source of bias in the estimate of the influence of the reference component; all bias works through the correlation between the reference component and the other two components. The center of each panel in the middle row sets the correlation between each program component and the omitted factor to zero, $\rho_{1\mu} = \rho_{2\mu} = \rho_{3\mu} = 0.00$, eliminating that source of bias. For these results, omitted-variable bias in Model 2 is the only source of bias.

Across all exhibits, results are favorable to the CAMIC method when program components are not correlated with the omitted factor.

Exhibit 4. Scenario Least Favorable to the CAMIC Method: Constant Correlation among Program Components of 0.70 ($\rho_{12} = \rho_{13} = \rho_{23} = 0.70$)

			pi3 = -0.50					pi3 = -0.25					pi3 = 0.25					pi3 = 0.50				
			rho3mu					rho3mu					rho3mu					rho3mu				
			-0.50	-0.25	0.00	0.25	0.50	-0.50	-0.25	0.00	0.25	0.50	-0.50	-0.25	0.00	0.25	0.50	-0.50	-0.25	0.00	0.25	0.50
rho1mu=-0.50	rho2mu	-0.50	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
		-0.25	🟥	🟥	🟩	🟩	🟩	🟥	🟥	🟩	🟩	🟩	🟥	🟥	🟥	🟩	🟩	🟥	🟥	🟥	🟥	🟩
		0.00	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
		0.25	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
		0.50	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
			-0.50	-0.25	0.00	0.25	0.50	-0.50	-0.25	0.00	0.25	0.50	-0.50	-0.25	0.00	0.25	0.50	-0.50	-0.25	0.00	0.25	0.50
rho1mu=-0.25	rho2mu	-0.50	🟥	🟥	🟩	🟩	🟩	🟥	🟥	🟩	🟩	🟩	🟥	🟥	🟥	🟩	🟩	🟥	🟥	🟥	🟥	🟩
		-0.25	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
		0.00	🟥	🟥	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
		0.25	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
		0.50	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
			-0.50	-0.25	0.00	0.25	0.50	-0.50	-0.25	0.00	0.25	0.50	-0.50	-0.25	0.00	0.25	0.50	-0.50	-0.25	0.00	0.25	0.50
rho1mu=0.00	rho2mu	-0.50	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
		-0.25	🟥	🟥	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
		0.00	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
		0.25	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟩	🟥	🟥	🟥	🟥
		0.50	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
			-0.50	-0.25	0.00	0.25	0.50	-0.50	-0.25	0.00	0.25	0.50	-0.50	-0.25	0.00	0.25	0.50	-0.50	-0.25	0.00	0.25	0.50
rho1mu=0.25	rho2mu	-0.50	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
		-0.25	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
		0.00	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟩	🟥	🟥	🟥	🟥
		0.25	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
		0.50	🟩	🟥	🟥	🟥	🟥	🟩	🟩	🟥	🟥	🟥	🟩	🟩	🟩	🟥	🟥	🟩	🟩	🟩	🟥	🟥
			-0.50	-0.25	0.00	0.25	0.50	-0.50	-0.25	0.00	0.25	0.50	-0.50	-0.25	0.00	0.25	0.50	-0.50	-0.25	0.00	0.25	0.50
rho1mu=0.50	rho2mu	-0.50	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
		-0.25	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
		0.00	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
		0.25	🟩	🟥	🟥	🟥	🟥	🟩	🟩	🟥	🟥	🟥	🟩	🟩	🟩	🟥	🟥	🟩	🟩	🟩	🟥	🟥
		0.50	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩

Of scenario results presented here, 29% are green and favorable to the CAMIC method.

Exhibit 5. Scenario Most Favorable to the CAMIC Method: Program Component #1 Very Highly Correlated with Program Component #2 ($\rho_{12} = 0.70$) and Highly Correlated with Program Component #3 ($\rho_{13} = 0.50$); Program Component #2 Somewhat Correlated with Program Component #3 ($\rho_{23} = 0.25$)

		pi3 = -0.50					pi3 = -0.25					pi3 = 0.25					pi3 = 0.50				
		rho3mu					rho3mu					rho3mu					rho3mu				
		-0.50	-0.25	0.00	0.25	0.50	-0.50	-0.25	0.00	0.25	0.50	-0.50	-0.25	0.00	0.25	0.50	-0.50	-0.25	0.00	0.25	0.50
rho1mu=-0.50	rho2mu -0.50	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
	-0.25	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
	0.00	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
	0.25	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
	0.50	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
		-0.50	-0.25	0.00	0.25	0.50	-0.50	-0.25	0.00	0.25	0.50	-0.50	-0.25	0.00	0.25	0.50	-0.50	-0.25	0.00	0.25	0.50
rho1mu=-0.25	rho2mu -0.50	🟩	🟩	🟩	🟥	🟥	🟩	🟩	🟩	🟥	🟥	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
	-0.25	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟩	🟩	🟥	🟥	🟥
	0.00	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
	0.25	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
	0.50	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
		-0.50	-0.25	0.00	0.25	0.50	-0.50	-0.25	0.00	0.25	0.50	-0.50	-0.25	0.00	0.25	0.50	-0.50	-0.25	0.00	0.25	0.50
rho1mu=0.00	rho2mu -0.50	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
	-0.25	🟩	🟩	🟩	🟩	🟥	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
	0.00	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
	0.25	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟥	🟩	🟩	🟩	🟩
	0.50	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
		-0.50	-0.25	0.00	0.25	0.50	-0.50	-0.25	0.00	0.25	0.50	-0.50	-0.25	0.00	0.25	0.50	-0.50	-0.25	0.00	0.25	0.50
rho1mu=0.25	rho2mu -0.50	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
	-0.25	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
	0.00	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
	0.25	🟥	🟥	🟥	🟩	🟩	🟥	🟥	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
	0.50	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟥	🟥	🟩	🟩	🟩	🟥	🟥	🟩	🟩	🟩
		-0.50	-0.25	0.00	0.25	0.50	-0.50	-0.25	0.00	0.25	0.50	-0.50	-0.25	0.00	0.25	0.50	-0.50	-0.25	0.00	0.25	0.50
rho1mu=0.50	rho2mu -0.50	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
	-0.25	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
	0.00	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
	0.25	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
	0.50	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥

Of scenario results presented here, 83% are green and favorable to the CAMIC method.

Exhibit 6. Scenario Most Typical of Favorability to the CAMIC Method: Program Component #1 Somewhat Correlated with Program Component #2 ($\rho_{12} = 0.25$) and Very Highly Correlated with Program Component #3 ($\rho_{13} = 0.70$); Program Component #2 Highly Correlated with Program Component #3 ($\rho_{23} = 0.50$)

		pi3 = -0.50					pi3 = -0.25					pi3 = 0.25					pi3 = 0.50				
		rho3mu					rho3mu					rho3mu					rho3mu				
		-0.50	-0.25	0.00	0.25	0.50	-0.50	-0.25	0.00	0.25	0.50	-0.50	-0.25	0.00	0.25	0.50	-0.50	-0.25	0.00	0.25	0.50
rho1mu=-0.50	rho2mu -0.50	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
	-0.25	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟥	🟩	🟩	🟩	🟩	🟥	🟩	🟩	🟩	🟩
	0.00	🟥	🟥	🟩	🟩	🟩	🟥	🟥	🟩	🟩	🟩	🟥	🟥	🟥	🟩	🟩	🟥	🟥	🟥	🟩	🟩
	0.25	🟥	🟥	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
	0.50	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
rho1mu=-0.25	rho2mu -0.50	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
	-0.25	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟥	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
	0.00	🟥	🟥	🟩	🟩	🟩	🟥	🟥	🟩	🟩	🟩	🟥	🟥	🟥	🟩	🟩	🟩	🟥	🟥	🟩	🟩
	0.25	🟥	🟥	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟥	🟩	🟥	🟥	🟥	🟥
	0.50	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟩	🟥	🟥	🟥	🟥
rho1mu=0.00	rho2mu -0.50	🟥	🟥	🟩	🟩	🟩	🟥	🟥	🟩	🟩	🟩	🟥	🟥	🟥	🟩	🟩	🟥	🟥	🟥	🟩	🟩
	-0.25	🟥	🟥	🟩	🟩	🟩	🟥	🟥	🟩	🟩	🟩	🟩	🟥	🟥	🟩	🟩	🟩	🟥	🟥	🟩	🟩
	0.00	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
	0.25	🟩	🟩	🟥	🟥	🟩	🟩	🟩	🟥	🟥	🟩	🟩	🟩	🟩	🟥	🟥	🟩	🟩	🟩	🟥	🟥
	0.50	🟩	🟩	🟥	🟥	🟥	🟩	🟩	🟥	🟥	🟥	🟩	🟩	🟩	🟥	🟥	🟩	🟩	🟩	🟥	🟥
rho1mu=0.25	rho2mu -0.50	🟥	🟥	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
	-0.25	🟥	🟥	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟩	🟥	🟥	🟥	🟥
	0.00	🟩	🟩	🟥	🟥	🟩	🟩	🟩	🟥	🟥	🟥	🟩	🟩	🟩	🟥	🟥	🟩	🟩	🟩	🟥	🟥
	0.25	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟥	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
	0.50	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
rho1mu=0.50	rho2mu -0.50	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
	-0.25	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟩	🟥	🟥	🟥	🟥
	0.00	🟩	🟩	🟥	🟥	🟥	🟩	🟩	🟥	🟥	🟥	🟩	🟩	🟩	🟥	🟥	🟩	🟩	🟩	🟥	🟥
	0.25	🟩	🟩	🟩	🟩	🟥	🟩	🟩	🟩	🟩	🟥	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
	0.50	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩

Of scenario results presented here, 54% are green and are favorable to the CAMIC method.

The scenarios presented in Exhibits 4 through 6 were selected to illustrate the range of findings across all scenarios. Among all scenarios investigated, the scenario presented in Exhibit 4 is least favorable to the CAMIC method with 29 percent of results favorable; the scenario presented in Exhibit 5 is the most favorable to the CAMIC method with 83 percent of results favorable. Across all findings for all scenarios, 53 percent of results were favorable to the CAMIC method. The scenario presented in Exhibit 6 comes closest to representing this average, with 54 percent of results favorable to the CAMIC method.

Taken together, Exhibits 4 through 6 demonstrate no obvious pattern of characteristics that produce results favorable to the CAMIC method. For example, the top panels of Exhibit 4 show results favorable to the CAMIC method in the top cells ($\rho_{1\mu} = \rho_{2\mu} = -0.50$), while these same cells are unfavorable to the CAMIC method in Exhibit 5. For each of these exhibits, the correlations among the program components, which are observable, are held constant; the correlation between each program component and the omitted factor and the coefficient of the third program component are allowed to vary. The lack of pattern across exhibits indicates that there is no pattern in unobservable characteristics associated with results that are more favorable to the CAMIC method. Without a pattern of favorable results, we are unable to identify conditions on the unobserved parameters that would yield results favorable to the CAMIC method. Doing so would have been useful in situations where suppositions about those parameters could help to enhance confidence in applying the method.

We next consider whether characteristics of observed parameters are associated with results more favorable to the CAMIC method.

Exhibits 7 through 9 summarize all scenarios investigated, presenting the proportion of results favorable to the CAMIC method for each of the specifications of the observed correlations among program components. Exhibit 7 displays results for all possible combinations of negative and positive signs for the correlations, holding the magnitude of the correlations constant at 0.25. For Scenarios 1, 4, 6, and 7, some 40 percent of findings are favorable to the CAMIC method. For Scenarios 2, 3, 5, and 8, some 66 percent of findings are favorable to the CAMIC method. The number of positive signs appears to determine the split: scenarios with an even number of negative correlations have 40 percent favorable, while scenarios with an odd number of negative correlations have 60 percent favorable. This is likely because an even number of negative terms multiplied together is positive. While sign appears to affect the findings to some extent, it does not appear to be a major source of variation in favorability to the CAMIC method.

Exhibit 7. Results for Scenarios, Focused on the Sign of the Correlations (Magnitudes Set to 0.25)

		Sign		Simulations Favorable to the CAMIC Method	
Scenario	ρ_{12}	ρ_{13}	ρ_{23}	N	Percent
1	+	+	+	200	40
2	+	-	+	330	66
3	+	+	-	330	66
4	+	-	-	200	40
5	-	+	+	330	66
6	-	-	+	200	40
7	-	+	-	200	40
8	-	-	-	330	66

Exhibit 8 displays results for scenarios with correlations increasing in magnitude. The correlations are equal to one another and increase from ($\rho_{12} = \rho_{13} = \rho_{23} = 0.10$) to ($\rho_{12} = \rho_{13} = \rho_{23} = 0.70$). The proportion of results favorable to the CAMIC method does not consistently increase or decrease as the magnitude increases. Rather, the proportion of scenarios favorable to the CAMIC method increases slightly from Scenario 9 to 12, then decreases from Scenario 12 to 16, increasing to Scenario 17 and then decreasing to Scenario 21. While the proportions do not vary over a wide range, the pattern of increases and decreases illustrates the nonlinearity of the relationship. If we used computational techniques to find the scenario with the maximum favorability to the CAMIC method, the approach to searching would be thrown off by the number of small peaks and valleys.

Exhibit 8. Results for Scenarios, Focused on the Magnitude of the Correlations

		Value		Simulations Favorable to the CAMIC method	
Scenario	ρ_{12}	ρ_{13}	ρ_{23}	N	Percent
9	0.10	0.10	0.10	192	38
10	0.15	0.15	0.15	192	38
11	0.20	0.20	0.20	192	38
12	0.25	0.25	0.25	200	40
13	0.30	0.30	0.30	192	38
14	0.35	0.35	0.35	188	38
15	0.40	0.40	0.40	180	36
16	0.45	0.45	0.45	176	35
17	0.50	0.50	0.50	188	38
18	0.55	0.55	0.55	156	31
19	0.60	0.60	0.60	152	30
20	0.65	0.65	0.65	144	29
21	0.70	0.70	0.70	144	29

Exhibit 9 presents results for scenarios that focus on the relative magnitude of the correlations. First, we consider all possible combinations of 0.25 and 0.50 as the value for the three correlations, including all correlations set to 0.25, all possible orderings of (0.50, 0.25, 0.25), all possible orderings of (0.50, 0.50, 0.25), and all correlations set to 0.5. Then, we consider all possible orderings of (0.70, 0.50, 0.25).

This exhibit presents the broadest range of findings, ranging from scenarios with 31 percent of simulations favorable to the CAMIC method to a scenario with 83 percent favorable to the CAMIC method. Comparing across scenarios, we see that a change in a single correlation can result in a large shift in findings or a minor shift, depending on the correlation and the starting point. For example, in scenarios 26 and 30, an increase in the correlation between program components #1 and #2 from 0.50 to 0.70 is associated with an increase in the proportion of simulations favorable to the CAMIC method from 43 to 83 percent. However, for scenarios 24 to 27, an increase of similar size in the correlation between program components #1 and #2 from 0.25 to 0.50 is associated with an increase in the proportion of simulations favorable to the CAMIC method from 40 to 43 percent.

Exhibit 9. Results for Scenarios, Focused on the Relative Magnitude of the Correlations

Scenario	Value			Simulations Favorable to the CAMIC method	
	ρ_{12}	ρ_{13}	ρ_{23}	N	Percent
All correlations set to 0.25					
22	0.25	0.25	0.25	200	40
All possible orders of (0.50,0.25,0.25)					
23	0.50	0.25	0.25	156	31
24	0.25	0.50	0.25	200	40
25	0.25	0.25	0.50	204	41
All possible orders of (0.50,0.50,0.25)					
26	0.50	0.50	0.25	214	43
27	0.50	0.25	0.50	274	55
28	0.25	0.50	0.50	238	48
All correlations set to 0.50					
29	0.50	0.50	0.50	188	38
All possible orders of (0.70,0.50,0.25)					
30	0.70	0.50	0.25	416	83
31	0.50	0.70	0.25	374	75
32	0.70	0.25	0.50	416	83
33	0.50	0.25	0.70	374	75
34	0.25	0.70	0.50	272	54
35	0.25	0.50	0.70	272	54

Although the specific proportions may be an artifact of the particular values of unobserved parameters we selected for the simulations, these findings fail to provide evidence supporting the intuition that the CAMIC method should help to reduce bias through better decisions over which variables to include in the model. These results demonstrate that the model that minimizes bias for the reference program component does not necessarily minimize bias for the focal program component.

These simulations calculate the bias in each model and compare them. By focusing on bias, the simulations ignore the role that variance might play in the CAMIC method. The estimates the CAMIC method works with differ from the true parameter value due to both bias and sampling variability in real data. The CAMIC method selects the model with the noisy measure of the reference component that is closest to the noisy, experimental measure of the reference component. Although the CAMIC method seeks to select the model that yields the least biased estimate of the reference component, noise in the estimates may result in selecting the wrong model. This limitation of the CAMIC method is not reflected in the simulations. However, the simulations show that even if the CAMIC method is able to consistently select the model that minimizes bias for the reference component, that model may or may not minimize bias in the target component.

5. Discussion and Conclusion

Large multi-site experiments with rich data on site and individual characteristics offer a unique opportunity to estimate the relationship between program components and implementation features and impact magnitude. However, if these program components and implementation features are not randomly assigned to individuals within sites (or to sites), then estimates of the relationship between them and impact magnitude may suffer from omitted-variable bias. In this paper, we provide a framework for leveraging the experimental evidence provided by a three-armed experiment to improve the non-experimental estimates of the impact contribution of program components and implementation features that occur naturally in the program world.

This inquiry parallels research considering “design replication studies,” also called “within-study comparison designs.” The goal of those studies (beginning with LaLonde, 1986) is to learn which approach to measuring program impact, subject to selection and other sources of bias using observational data, best replicates an experimental finding for the same impact quantity. A series of such studies has begun to point evaluators toward the conditions that yield more-reliable non-experimental findings than other options do (see Cook et al., 2008; Glazerman et al., 2003), when prior to LaLonde (1986) no empirical guide along that path was known. In a similar way, CAMIC method-based tests of cross-site impact attribution specifications may yield consistent results with enough replications—especially if multiple program components and/or implementation features can be randomized into enhanced treatment arms within a single multi-site study. Certainly as the state of the science of within-study comparisons improves, we should gain knowledge regarding the calibration of non-experimental to experimental results.

That said, the potential benefit of using the CAMIC method is not simply to compare a non-experimental result with an experimental benchmark, but instead to use that comparison to improve other non-experimental results. As such, the method advances design replication studies to extended applications. The evaluation challenge at hand is that of selection bias that occurs at the site level in multi-site evaluations with individual-level random assignment and results in omitted-variable bias in the estimates of the influence of selected program components and implementation features on overall impacts.

The HPOG Impact Study may provide an opportunity to empirically gauge the utility of the CAMIC method from real-world data. This is because, as discussed earlier, HPOG randomizes three program components to additional treatment arms as enhancements to basic programs, each in a different set of sites. Those same program components—facilitated peer support groups, emergency assistance for specific needs, and noncash incentives that encourage desirable program outputs and outcomes—also exist naturally as part of certain standard programs.

This presents an opportunity to test the CAMIC method’s success across different enhancement components. For example, researchers can check whether the CAMIC method model calibrated to replicate the experimental finding on peer support also gets the emergency assistance estimate right non-experimentally—that is, it produces an emergency assistance estimate that aligns with experimental evidence on that component’s effect. Reversing this, researchers also can check whether the CAMIC method model calibrated to replicate the experimental finding on emergency assistance also accurately estimates the influence of peer support non-experimentally—that is, it produces a peer support estimate that aligns with experimental evidence on that component's effect. This cross-calibration, if you will, and associated impact estimation can be further enhanced by using the third component randomly assigned as an enhancement, noncash incentives. As a result, HPOG will have experimental evidence on the contribution of *three* program components with which to gauge the bias of non-experimental estimates of those same contributions, in its own within-study comparison. Moreover, HPOG has the opportunity to use each of these three "checks" of the success of the CAMIC method in the HPOG application to decide how much credence to give findings from—and to further attempt to improve findings from—non-experimental CAMIC method-based estimates of the contributions of *other* program components and implementation features to program impact.

Only a real-world application of this sort will be able to provide needed additional insights on the CAMIC method's performance and utility. Having one lined up for the HPOG Impact Study is encouraging. We believe that HPOG's "multiple-enhancement three-arm multi-site experimental" design provides a model for future social program evaluations aiming to focus on the relative effectiveness of selected program components in a rigorous way.

While waiting for the HPOG outcome data to become available to the study team, this paper reports on simulations aimed at exploring the potential for the CAMIC method to succeed in these ways under various scenarios. In addition to the simulation results reported here, a prior simulation was conducted to explore the CAMIC method's potential reliability in other respects. Neither of these two simulation studies has pointed to clear conditions for the CAMIC method's success.

The simulation exercises explored a range of scenarios using a simplified framework. The framework was developed to represent the key relationships in the analysis, and the scenarios were developed to systematically explore the range of possibilities. However, it remains possible that we would draw a different set of conclusions had we made different simplifying assumptions or explored a different set of scenarios. Future work might consider three extensions. First, we support efforts to determine the conditions under which the CAMIC method is an improvement over standard practice in producing minimally biased non-experimental estimates of the contribution of selected program components or implementation features to program impacts. Even if improvements cannot occur in the majority of cases (as our existing simulation investigations suggest, with results favoring the CAMIC method in between 29 and 83 percent of the examined circumstances), at least (1) the limits on the CAMIC method's utility could be better understood, and (2) there is a possibility to learn what characterizes the most favorable (or least favorable) conditions for its application.

Second, we suggest exploring when multi-level modeling that is not in a position to use the CAMIC method as a tool can and cannot eliminate bias caused by omitted factors. We hypothesize that this kind of bias will remain an issue under many, perhaps increasingly complex, analytic strategies given the complex, non-linear dependencies among site-level variables. Knowledge of this sort would be of value to a field that is increasingly using multi-site evaluations with individual-level randomization to measure the contribution of program components and implementation features to program impacts in a multi-faceted program.

Third—and we believe most valuable—would be future work that applies the CAMIC method to a variety of actual randomized impact evaluations in a range of settings. This will be facilitated by the increasing use of multi-site experimental evaluation designs in general, offering more opportunities for creating a third experimental arm to measure the effect of individual elements within multi-faceted programs. HPOG will serve as the first such opportunity, and the first applied test of the CAMIC method. We hope the analysis team for that research will emerge with added insights and useful lessons for future evaluation design and analytic work to push forward innovative methods such as the CAMIC method that can help improve the capabilities of social program impact evaluators by expanding their toolkit.

6. References

Bloom, H.S., Hill, C.J., & Riccio, J.A. (2001). *Modeling the performance of welfare-to-work programs: The effects of program management and services, economic environment, and client characteristics.* New York, NY: Manpower Demonstration Research Corporation.

Bloom, H.S., Hill, C.J., & Riccio, J.A. (2003). Linking program implementation and effectiveness: Lessons from a pooled sample of welfare-to-work experiments. *Journal of Policy Analysis and Management*, 22, 551–575.

Bryk, A., & Raudenbush, S. (1992). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage.

Cook, T.D., Shadish, W.R., & Wong, V.C. (2008). Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis and Management*, 27, 725–750.

Dorsett, R., & Robins, P.K. (2014). A multilevel analysis of the impacts of services provided by the U.K. Employment Retention and Advancement demonstration. *Evaluation Review*, 37, 63–108. doi: 10.1177/0193841X13517383

Glazerman, S., Levy, D.M., & Myers, D. (2003). Nonexperimental versus experimental estimates of earnings impacts. *The Annals of the American Academy of Political and Social Science*, 589, 63-93.

Godfrey, E.B., & Yoshikawa, H. (2012). Caseworker recipient interaction: Welfare office differences, economic trajectories, and child outcomes. *Child Development*, 83, 382–398.

Greenberg, D., Meyer, R., & Wiseman, M. (1994). Multi-site employment and training evaluations: A tale of three studies. *Industrial and Labor Relations Review*, 47, 679–691.

LaLonde, R. (1986). Evaluating the econometric evaluations of training with experimental data. *American Economic Review*, 76, 604–620.

Steiner, P.M., & Kim, Y. (2015). “Omitted variable bias: Bias amplification and cancellation of offsetting biases.” Paper presented at the Fall Conference of the Association for Public Policy Analysis and Management, Miami, FL, November 14.

Wooldridge, J.M. (2002). *Econometric analysis of cross section and panel data.* Cambridge, MA: The MIT Press.

Appendix A. Source of Omitted-Variable Bias in Non-Experimental

Estimating the model described in Section 2 will produce estimates of the π_m and φ_g parameters to describe the relationship between program components and implementation features and treatment impact. Omitted-variable bias arises in this analysis if a site-level factor F_j exists that (a) influences impact magnitude, (b) is not included as a right-hand-side variable in Equation (2) or Equation (4) of the Section 2 model, and (c) correlates with one or more of the components or features included as right-hand-side variables in the model specification. Assuming that F_j has a linear influence—and that there is no interaction between it and the included components or features when determining impact magnitude—the correct specification of Equation (2) from the main text is as follows:

$$\beta_j = \beta_0 + \sum_m \pi_m P_{mj} + \sum_g \varphi_g I_{gj} + \sum_d \tau_d PC_{dj} + \sum_q \zeta_q LC_{qj} + \lambda F_j + \mu_j \qquad \text{(Eq. 2}'\text{)}$$

where:

λ = influence of the omitted factor on impact magnitude; and

F_j = amount of omitted factor in site j.

Here, F_j , if unmeasured, becomes an omitted confounder in the analysis of the determinants of impact magnitude described in the main text and revisited below.

To see how bias arises, plug Equation (2′) and Equation (3) into Equation (1) to get:

$$Y_{ji} = \alpha_0 + \sum_q \kappa_q LC_{qj} + \beta_0 T_{ji} + \sum_m \pi_m P_{mj} T_{ji} + \sum_g \varphi_g I_{gj} T_{ji} + \sum_d \tau_d PC_{dj} T_{ji} +$$

$$\sum_q \zeta_q LC_{qj} T_{ji} + \lambda F_j T_{ji} + \sum_c \delta_c IC_{cji} + \sum_c \gamma_c IC_{cji} T_{ji} + \{ v_j + \mu_j T_{ji} + \varepsilon_{ji} \} \qquad \text{(Eq. 4}'\text{)}$$

When this equation is estimated with maximum likelihood methods (Bryk & Raudenbush, 1992) with $F_j T_{ji}$ omitted as in Equation (4), the probability limits of the resulting estimators of the π_m and φ_g coefficients, $\hat{\pi}_m$ and $\hat{\varphi}_g$, do not equal π_m and φ_g. That is, the estimates of the contribution of every program component and implementation feature to impact magnitude are asymptotically biased. As an example, it can be shown that the asymptotic expectation of $\hat{\pi}_m$ in the special case where F_j correlates with program component P_{rj} but not with any of the other right-hand-side variables in Equation (4′), is

$$plim(\hat{\pi}_r) = \pi_r + \lambda \frac{\mathrm{Cov}\,(P_{rj}, F_j)}{\mathrm{Var}(P_{rj})}$$

(see Wooldridge, 2002, pp. 61-62). This equation aligns with the usual econometric formula for omitted-variable bias, although it gives the relationship at the limit as sample size goes to infinity. The bias of $\hat{\pi}_r$ can be derived from this probability limit as

$$bias(\hat{\pi}_r) = plim(\hat{\pi}_r) - \pi_r = \lambda \frac{\mathrm{Cov}\,(P_{rj}, F_j)}{\mathrm{Var}(P_{rj})},$$

because $Cov(P_{rj}, F_j) \neq 0$, $\hat{\pi}_r$ from the mis-specified Equation (4) model is a biased estimate of π_r even in very large samples, with bias that approaches $\lambda \frac{\text{Cov}\left(P_{rj}, F_j\right)}{\text{Var}(P_{rj})}$ as sample size goes to infinity. The same argument can be made for all of $\hat{\pi}_1$,..., $\hat{\pi}_M$ and $\hat{\varphi}_{g1}$,..., $\hat{\varphi}_{gG}$.

The more general case where more than one right-hand-side variable in Equation (4′) correlates with F_j produces estimates for which

$$plim(\hat{\pi}_1) = \pi_1 + Z_1$$

$$plim(\hat{\pi}_M) = \pi_M + Z_M$$

$$plim(\hat{\varphi}_1) = \varphi_1 + W_1$$

$$plim(\hat{\varphi}_G\ ) = \varphi_G + W_{G,}$$

where Z_1,..., Z_M and W_1,..., W_M are all non-zero but have complex mathematical expressions not provided by Wooldridge. Regardless of their forms,

$$bias(\hat{\pi}_m) = plim(\hat{\pi}_m) - \pi_m = Z_m \neq 0 \text{ for } m = 1,..., M, \text{ and}$$

$$bias(\hat{\varphi}_g) = plim(\hat{\varphi}_g\ ) - \varphi_g = W_g \neq 0 \text{ for } g = 1,..., G.$$

Appendix B. Full Specification of the CAMIC Method Step 4 Model to

This appendix shows the modifications to Equation (4) in Section 2 that are necessary to estimate that model using data on all three experimental arms—standard treatment, enhanced treatment, and control group—in all evaluation sties. Specifically, the final step in the CAMIC method is estimated on the following model:

Level One: Individuals

$$Y_{ji} = \alpha_j + \beta_j TE_{ji} + \pi_{ej} E_{ji} + \sum_c \delta_c IC_{cji} + \sum_c \gamma_c IC_{cji}\, TE_{ji} + \varepsilon_{ji} \qquad \text{(Eq. 10)}$$

Level Two: Sites

$$\beta_j = \beta_0 + \sum_m \pi_m\, P_{mj} + \sum_g \varphi_g I_{gj} + \sum_d \tau_d PC_{dj} + \sum_q \zeta_q LC_{qj} + \mu_j \qquad \text{(Eq. 11)}$$

$$\pi_{ej} \ = \pi_e + \omega_j \qquad \text{(Eq. 12)}$$

and:

$$\alpha_j = \alpha_0 + \sum_q \kappa_q LC_{qj} + \ \upsilon_j \qquad \text{(Eq. 13)}$$

Combining the elements of the above two-level model produces the following:

$$\begin{aligned} Y_{ji} = {} & \alpha_0 + \sum_q \kappa_q LC_{qj} + \beta_0 TE_{ji} + \sum_m \pi_m\, P_{mj} TE_{ji} + \sum_g \varphi_g I_{gj} TE_{ji} + \\ & \sum_d \tau_d PC_{dj} TE_{ji} + \sum_q \zeta_q LC_{qj} TE_{ji} + \pi_e E_{ji} + \sum_c \delta_c IC_{cji} + \sum_c \gamma_c IC_{cji}\, TE_{ji} + \\ & \{\upsilon_j + \mu_j T_{ji} + \omega_j E_{ji} + \ \varepsilon_{ji}\} \end{aligned} \qquad \text{(Eq. 14)}$$

We can simplify Equation (14) by combining $\pi_e P_{ej} TE_{ji}$ from the third summation term and $\pi_e E_{ji}$ to get the following:

$$\begin{aligned} Y_{ji} = {} & \alpha_0 + \sum_q \kappa_q LC_{qj} + \beta_0 TE_{ji} + \sum_{m \neq e} \pi_m\, P_{mj} TE_{ji} + \sum_g \varphi_g I_{gj} TE_{ji} + \\ & \sum_d \tau_d PC_{dj} TE_{ji} + \sum_q \zeta_q LC_{qj} TE_{ji} + \pi_e (P_{ej} TE_{ji} + E_{ji}) + \sum_c \delta_c IC_{cji} + \\ & \sum_c \gamma_c IC_{cji}\, TE_{ji} + \{\upsilon_j + \mu_j T_{ji} + \omega_j E_{ji} + \varepsilon_{ji}\} \end{aligned} \qquad \text{(Eq. 15)}$$

Notation definitions appear in Exhibit 1. Note that Equation (10) is a modified version of Equation (1) where T_{ji}—the indicator for whether a given study member is assigned to the standard treatment—is replaced by TE_{ji}—the indicator for whether a given study member is assigned to either the standard treatment or the enhanced treatment. Additionally, Equation (10) adds to Equation (1) a further randomly assigned program component in the enhancement arm, component m. Next, π_e, the coefficient on the indicator for randomization to that arm, E_{ji}, gets its own expanded expression and error term in Equation (12) just as has β_j, the coefficient on the other randomization indicator in Equation (2)—when that indicator was T_{ji}—and in Equation (11)—where that indicator is now TE_{ji}.

Appendix C. Deriving Expressions for Bias in Simplified Framework

To explore bias in the simplified framework, we move to matrix notation by stacking observations to consider the properties of the estimator.

$$\underbrace{\begin{bmatrix}\Delta_1\\ \Delta_2\\ \vdots\\ \Delta_J\end{bmatrix}}_{J\times 1} = \underbrace{\begin{bmatrix}1 & P_{11} & P_{21} & P_{31}\\ 1 & P_{12} & P_{22} & P_{32}\\ \vdots & \vdots & \vdots & \vdots\\ 1 & P_{1J} & P_{2J} & P_{3J}\end{bmatrix}}_{J\times 4}\underbrace{\begin{bmatrix}\pi_0\\ \pi_1\\ \pi_2\\ \pi_3\end{bmatrix}}_{4\times 1} + \underbrace{\begin{bmatrix}\mu_1\\ \mu_2\\ \vdots\\ \mu_J\end{bmatrix}}_{J\times 1} + \underbrace{\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \vdots\\ \varepsilon_J\end{bmatrix}}_{J\times 1}$$

$$\Delta = P\pi + \mu + \varepsilon$$

Running ordinary least squares (OLS) yields:

$$\hat{\pi} = (P'P)^{-1}P'\Delta$$

With expected value:

$$E\{\hat{\pi}\} = E\{(P'P)^{-1}P'\Delta\}$$

$$= E\{(P'P)^{-1}P'(P\pi + \mu + \varepsilon)\}$$

$$= \ E\{(P'P)^{-1}P'P\pi\} + E\{(P'P)^{-1}P'\mu\} + E\{(P'P)^{-1}P'\varepsilon\}$$

$$= \pi + E\{(P'P)^{-1}P'\mu\}$$

Therefore, bias is given by:

$$E\{\hat{\pi}\} - \pi = E\{(P'P)^{-1}P'\mu\}$$

Note that the standard approach is to take this expectation conditional on the independent variables. However, that approach misrepresents the structure of uncertainty. It seems that the unobserved factor is likely realized either before the observed program components (if it is a characteristic of leadership) or at the same time as the observed program components (if it is an omitted program component). In either case, it does not seem useful to think of a world in which a program with a fixed set of components fields the intervention an infinite number of times with different values of the unobserved factor. Instead, it makes sense to conceive of a world in which programs simultaneously select observed and unobserved program components based on some distribution associating the two. So that is the world in which we explore bias.

If the unobserved factor was conditionally independent of the program components ($E\{\mu|P\} = 0$), then we could simplify this term as follows:

$$E\{(P'P)^{-1}P'\mu\} = \ E_P\left\{E_{\mu|P}\{(P'P)^{-1}P'\mu|P\}\right\} = E_P\left\{(P'P)^{-1}P'E_{\mu|P}\{\mu|P\}\right\} = E_P\left\{(P'P)^{-1}P'\vec{0}\right\}$$
$$= 0$$

However, in that case there is no bias and no reason to go forward with this method.

Given that the unobserved factor is related to one or more of the observed program components, the expression for the bias cannot be simplified further. We cannot distribute the expectation across the terms of the product, and the expectation of an inverse is not the inverse of the expectation. However, it is worth thinking through the interpretation of these terms to gain some intuition.

Bias in Simplified Framework

The term $P'P$ includes estimates of the variance and covariance of the observed program components.

$$P'P = \begin{bmatrix} J & \sum_j P_{1j} & \sum_j P_{2j} & \sum_j P_{3j} \\ \sum_j P_{1j} & \sum_j P_{1j}^2 & \sum_j P_{1j}P_{2j} & \sum_j P_{1j}\, P_{3j} \\ \sum_j P_{2j} & \sum_j P_{1j}P_{2j} & \sum_j P_{2j}^2 & \sum_j P_{2j}\, P_{3j} \\ \sum_j P_{3j} & \sum_j P_{1j}\, P_{3j} & \sum_j P_{2j}\, P_{3j} & \sum_j P_{3j}^2 \end{bmatrix}$$

$$= J \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \widehat{Var}(P_1) & \widehat{Cov}(P_1, P_2) & \widehat{Cov}(P_1, P_3) \\ 0 & \widehat{Cov}(P_1, P_2) & \widehat{Var}(P_2) & \widehat{Cov}(P_2, P_3) \\ 0 & \widehat{Cov}(P_1, P_3) & \widehat{Cov}(P_2, P_3) & \widehat{Var}(P_3) \end{bmatrix}$$

The term $P'\mu$ includes the realized covariance between the program components and the unobserved factor that we would obtain if we observed the factor.

$$P'\mu = \begin{bmatrix} \sum_j \mu_j \\ \sum_j \mu_j P_{1j} \\ \sum_j \mu_j P_{2j} \\ \sum_j \mu_j P_{3j} \end{bmatrix} = J \begin{bmatrix} 0 \\ \widehat{Cov}(\mu, P_1) \\ \widehat{Cov}(\mu, P_2) \\ \widehat{Cov}(\mu, P_3) \end{bmatrix}$$

Multiplying out $(P'P)^{-1}P'\mu$ yields the following (incredibly messy) expressions:

$$(P'P)^{-1} = \frac{1}{J\det(P'P)} * \begin{bmatrix} \det(P'P) & 0 & 0 & 0 \\ 0 & \widehat{Var}(P_2)\widehat{Var}(P_3) - \widehat{Cov}(P_2,P_3)^2 & \widehat{Cov}(P_1,P_3)\widehat{Cov}(P_2,P_3) - \widehat{Cov}(P_1,P_2)\widehat{Var}(P_3) & \widehat{Cov}(P_1,P_2)\widehat{Cov}(P_2,P_3) - \widehat{Var}(P_2)\widehat{Cov}(P_1,P_3) \\ 0 & \widehat{Cov}(P_1,P_3)\widehat{Cov}(P_2,P_3) - \widehat{Cov}(P_1,P_2)\widehat{Var}(P_3) & \widehat{Var}(P_1)\widehat{Var}(P_3) - \widehat{Cov}(P_1,P_3)^2 & \widehat{Cov}(P_1,P_2)\widehat{Cov}(P_1,P_3) - \widehat{Var}(P_1)\widehat{Cov}(P_2,P_3) \\ 0 & \widehat{Cov}(P_1,P_2)\widehat{Cov}(P_2,P_3) - \widehat{Var}(P_2)\widehat{Cov}(P_1,P_3) & \widehat{Cov}(P_1,P_2)\widehat{Cov}(P_1,P_3) - \widehat{Var}(P_1)Cov(P_2,P_3) & \widehat{Var}(P_1)\widehat{Var}(P_3) - \widehat{Cov}(P_1,P_3)^2 \end{bmatrix},$$

where $\det(P'P) = \widehat{Var}(P_1)\widehat{Var}(P_2)\widehat{Var}(P_3) - \widehat{Var}(P_1)\widehat{Cov}(P_2,P_3)^2 - \widehat{Var}(P_3)\widehat{Cov}(P_1,P_2)^2 - \widehat{Var}(P_2)\widehat{Cov}(P_1,P_3)^2 + 2\widehat{Cov}(P_1,P_2)\widehat{Cov}(P_1,P_3)\widehat{Cov}(P_2,P_3)$

$$(P'P)^{-1}P'\mu = \begin{bmatrix} 0 \\ \dfrac{\left(\widehat{Var}(P_2)\widehat{Var}(P_3) - \widehat{Cov}(P_2,P_3)^2\right)\widehat{Cov}(\mu,P_1) + \left(\widehat{Cov}(P_1,P_3)\widehat{Cov}(P_2,P_3) - \widehat{Cov}(P_1,P_2)\widehat{Var}(P_3)\right)\widehat{Cov}(\mu,P_2) + \left(\widehat{Cov}(P_1,P_2)\widehat{Cov}(P_2,P_3) - \widehat{Var}(P_2)\widehat{Cov}(P_1,P_3)\right)\widehat{Cov}(\mu,P_3)}{\widehat{Var}(P_1)\widehat{Var}(P_2)\widehat{Var}(P_3) - \widehat{Var}(P_1)\widehat{Cov}(P_2,P_3)^2 - \widehat{Var}(P_3)\widehat{Cov}(P_1,P_2)^2 - \widehat{Var}(P_2)\widehat{Cov}(P_1,P_3)^2 + 2\widehat{Cov}(P_1,P_2)\widehat{Cov}(P_1,P_3)\widehat{Cov}(P_2,P_3)} \\ \dfrac{\left(\widehat{Cov}(P_1,P_3)\widehat{Cov}(P_2,P_3) - \widehat{Cov}(P_1,P_2)\widehat{Var}(P_3)\right)\widehat{Cov}(\mu,P_1) + \left(\widehat{Var}(P_1)\widehat{Var}(P_3) - \widehat{Cov}(P_1,P_3)^2\right)\widehat{Cov}(\mu,P_2) + \left(\widehat{Cov}(P_1,P_2)\widehat{Cov}(P_1,P_3) - \widehat{Var}(P_1)\widehat{Cov}(P_2,P_3)\right)\widehat{Cov}(\mu,P_3)}{\widehat{Var}(P_1)\widehat{Var}(P_2)\widehat{Var}(P_3) - \widehat{Var}(P_1)\widehat{Cov}(P_2,P_3)^2 - \widehat{Var}(P_3)\widehat{Cov}(P_1,P_2)^2 - \widehat{Var}(P_2)\widehat{Cov}(P_1,P_3)^2 + 2\widehat{Cov}(P_1,P_2)\widehat{Cov}(P_1,P_3)\widehat{Cov}(P_2,P_3)} \\ \dfrac{\left(\widehat{Cov}(P_1,P_2)\widehat{Cov}(P_2,P_3) - \widehat{Var}(P_2)\widehat{Cov}(P_1,P_3)\right)\widehat{Cov}(\mu,P_1) + \left(\widehat{Cov}(P_1,P_2)\widehat{Cov}(P_1,P_3) - \widehat{Var}(P_1)Cov(P_2,P_3)\right)\widehat{Cov}(\mu,P_2) + \left(\widehat{Var}(P_1)\widehat{Var}(P_3) - \widehat{Cov}(P_1,P_3)^2\right)\widehat{Cov}(\mu,P_3)}{\widehat{Var}(P_1)\widehat{Var}(P_2)\widehat{Var}(P_3) - \widehat{Var}(P_1)\widehat{Cov}(P_2,P_3)^2 - \widehat{Var}(P_3)\widehat{Cov}(P_1,P_2)^2 - \widehat{Var}(P_2)\widehat{Cov}(P_1,P_3)^2 + 2\widehat{Cov}(P_1,P_2)\widehat{Cov}(P_1,P_3)\widehat{Cov}(P_2,P_3)} \end{bmatrix}$$

What can we learn from these equations?

- The bias in our estimate of program component π_1 is the statistical expectation of a non-linear function of all terms:
  - Observed variance in program components: $\widehat{Var}(P_1), \widehat{Var}(P_2), \widehat{Var}(P_3)$
  - Observed covariance between components: $\widehat{Cov}(P_1, P_2), \widehat{Cov}(P_1, P_3), \widehat{Cov}(P_2, P_3)$
  - Realized covariance between program components and the unobserved factor: $\widehat{Cov}(\mu, P_1), \widehat{Cov}(\mu, P_2), \widehat{Cov}(\mu, P_3)$
- Although it is tempting to simply substitute the true variances and covariances into the expression above, that does not align with the properties of the expectation operator. We cannot distribute the expectation across the terms of the product, and the expectation of an inverse is not the inverse of the expectation.
- However, it is reasonable to hypothesize that all the true variances and covariances affect the expected bias in each of the program components.
- The bias does not depend on the actual relationship between the program components and the impact of the intervention.
- The bias does not depend on the variance of the unobserved factor or of the error term (though these will enter the variance of the estimator).

Bias in Model 2

The derivation of the bias in Model 2 is analogous to the derivation for Model 1. Let π^2 be the vector of true parameter values that can be estimated using Model 2.

$$E\{\hat{\pi}^2\} - \pi^2 = E\left\{\begin{bmatrix} J & \sum_j P_{1j} & \sum_j P_{2j} \\ \sum_j P_{1j} & \sum_j P_{1j}^2 & \sum_j P_{1j}P_{2j} \\ \sum_j P_{2j} & \sum_j P_{1j}P_{2j} & \sum_j P_{2j}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_j (\pi_3 P_{3j} + \mu_j) \\ \sum_j (\pi_3 P_{3j} + \mu_j) P_{1j} \\ \sum_j (\pi_2 P_{2j} + \mu_j) P_{2j} \end{bmatrix}\right\}$$

$$= E\left\{\begin{bmatrix} 1 & 0 & 0 \\ 0 & \widehat{Var}(P_1) & \widehat{Cov}(P_1, P_2) \\ 0 & \widehat{Cov}(P_1, P_2) & \widehat{Var}(P_2) \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ \pi_3 \widehat{Cov}(P_1, P_3) + \widehat{Cov}(\mu, P_1) \\ \pi_3 \widehat{Cov}(P_1, P_2) + \widehat{Cov}(\mu, P_2) \end{bmatrix}\right\}$$

$$= E\left\{\begin{bmatrix} 1 & 0 & 0 \\ 0 & \dfrac{\widehat{Var}(P_2)}{\widehat{Var}(P_1)\widehat{Var}(P_2) - \widehat{Cov}(P_1, P_2)^2} & \dfrac{-\widehat{Cov}(P_1, P_2)}{\widehat{Var}(P_1)\widehat{Var}(P_2) - \widehat{Cov}(P_1, P_2)^2} \\ 0 & \dfrac{-\widehat{Cov}(P_1, P_2)}{\widehat{Var}(P_1)\widehat{Var}(P_2) - \widehat{Cov}(P_1, P_2)^2} & \dfrac{\widehat{Var}(P_1)}{\widehat{Var}(P_1)\widehat{Var}(P_2) - \widehat{Cov}(P_1, P_2)^2} \end{bmatrix}\right..$$

$$\left.\left.\cdot\begin{bmatrix} 0 \\ \pi_3\widehat{Cov}(P_1,P_3)+\widehat{Cov}(\mu,P_1) \\ \pi_3\widehat{Cov}(P_2,P_3)+\widehat{Cov}(\mu,P_2)\end{bmatrix}\right)\right\}$$

$$= E\left\{\left(\begin{bmatrix} 0 \\ \dfrac{\widehat{Var}(P_2)\big(\pi_3\widehat{Cov}(P_1,P_3)+\widehat{Cov}(\mu,P_1)\big)-\widehat{Cov}(P_1,P_2)\big(\pi_3\widehat{Cov}(P_2,P_3)+\widehat{Cov}(\mu,P_2)\big)}{\widehat{Var}(P_1)\widehat{Var}(P_2)-\widehat{Cov}(P_1,P_2)^2} \\ \dfrac{\widehat{Var}(P_1)\big(\pi_3\widehat{Cov}(P_2,P_3)+\widehat{Cov}(\mu,P_2)\big)-\widehat{Cov}(P_1,P_2)\big(\pi_3\widehat{Cov}(P_1,P_3)+\widehat{Cov}(\mu,P_1)\big)}{\widehat{Var}(P_1)\widehat{Var}(P_2)-\widehat{Cov}(P_1,P_2)^2}\end{bmatrix}\right)\right\}$$

Appendix D. Full Specifications for All Scenarios

This Appendix details the elements of the simulation scenarios in terms of their focus of exploration, and the input values of each of the variables, as shown in Exhibit D.1.

Exhibit D.1. Details of Scenarios Examined to Analyze the CAMIC Method's Potential

Scenario	Focus of Exploration	ρ_{12}	ρ_{13}	ρ_{23}
1	Sign	0.25	0.25	0.25
2	Sign	0.25	-0.25	0.25
3	Sign	0.25	0.25	-0.25
4	Sign	0.25	-0.25	-0.25
5	Sign	-0.25	0.25	0.25
6	Sign	-0.25	-0.25	0.25
7	Sign	-0.25	0.25	-0.25
8	Sign	-0.25	-0.25	-0.25
9	Magnitude	0.10	0.10	0.10
10	Magnitude	0.15	0.15	0.15
11	Magnitude	0.20	0.20	0.20
12	Magnitude	0.25	0.25	0.25
13	Magnitude	0.30	0.30	0.30
14	Magnitude	0.35	0.35	0.35
15	Magnitude	0.40	0.40	0.40
16	Magnitude	0.45	0.45	0.45
17	Magnitude	0.50	0.50	0.50
18	Magnitude	0.55	0.55	0.55
19	Magnitude	0.60	0.60	0.60
20	Magnitude	0.65	0.65	0.65
21	Magnitude	0.70	0.70	0.70
22	Relative Magnitude	0.25	0.25	0.25
23	Relative Magnitude	0.50	0.25	0.25
24	Relative Magnitude	0.25	0.50	0.25
25	Relative Magnitude	0.25	0.25	0.50
26	Relative Magnitude	0.50	0.50	0.25
27	Relative Magnitude	0.50	0.25	0.50
28	Relative Magnitude	0.25	0.50	0.50
29	Relative Magnitude	0.50	0.50	0.50
30	Relative Magnitude	0.70	0.50	0.25
31	Relative Magnitude	0.50	0.70	0.25
32	Relative Magnitude	0.70	0.25	0.50
33	Relative Magnitude	0.50	0.25	0.70
34	Relative Magnitude	0.25	0.70	0.50
35	Relative Magnitude	0.25	0.50	0.70