Hypothesis Testing in the Absence of Microdata*

Vedant Vohra

August 21, 2024

Abstract

Economists are often interested in functions of multiple causal effects. A leading example is when evaluating the cost-effectiveness of a policy. In such settings, the benefits and costs are proxied by multiple causal effects and aggregated into a scalar measure of cost-effectiveness. Oftentimes, access to the microdata underlying the estimates is infeasible; only the published estimates and their corresponding standard errors are available for post-hoc analysis. We provide a method to conduct inference on functions of causal effects when the only information available is the point estimates and their corresponding standard errors. We apply these methods to conduct inference on the Marginal Value of Public Funds (MVPF) for 8 different policies, and show that even in the absence of any microdata, our method can be used to conduct valid but possibly conservative inference on the MVPF.

JEL Numbers: C12, C21, C30.

*Vohra: UC San Diego. I am grateful to Julie Cullen, Graham Elliott, Itzik Fadlon, Stefan Faridani, Xinwei Ma, Craig McIntosh, and Kaspar Wuthrich for helpful comments.

1 Introduction

Economists are often interested in functions of multiple causal effects. For instance, one goal of policy evaluations is determining the policy's cost-effectiveness: does a dollar of expenditure on the policy yield more than a dollar of benefits? A policy evaluation of a cash transfer program might estimate its causal effect on a range of different outcomes and seek to determine whether a government should fund it. The benefits of a cash transfer might include improved health, education, and financial well-being; the costs might include potential labor market disincentives. These causal effects would need to be aggregated into a scalar measure to obtain an estimate for the cost-effectiveness of the program. When the target parameter of interest (e.g., cost-effectiveness of the cash transfer program) is a scalar-valued non-linear function of multiple causal effects, how can we conduct inference on this parameter? If we had access to the microdata underlying each causal effect (for instance, the data from an RCT evaluating the causal effects of a cash transfer), we could easily estimate the correlation across each causal effect and estimate the variance of the function. However, in many settings of interest, estimating the full correlation structure across causal effects is infeasible. In this paper, we study inference on functions of multiple causal effects, without knowledge about the correlation structure across these causal effects.

To help motivate the question, we focus on the problem of conducting inference on the Marginal Value of Public Funds (MVPF) for a given policy (Hendren and Sprung-Keyser, 2020). The MVPF is a unified metric for evaluating the welfare consequences of government policies. It is calculated as a non-linear function of multiple causal effects: the benefits that a policy provides to its recipients divided by the policy's net cost to the government. Hendren and Sprung-Keyser (2020) use causal estimates available in existing studies to construct the MVPF for over a hundred policies. The only information available to estimate the MVPF and its variance are the causal effects and their corresponding standard errors reported in the original study. The microdata underlying these causal estimates is inaccessible to Hendren and Sprung-Keyser (2020) for ex-post analysis. The fundamental challenge is that the variance of the MVPF depends not only on the standard errors of the causal effects but also on the correlation across the causal effects. These correlation are rarely reported in papers and cannot be estimated in the absence of microdata. The goal of this paper is to provide a simple inference procedure that yields valid confidence intervals on functions of causal effect without knowledge about the correlation across these causal effects.

Estimates of the correlation structure across causal effects might be difficult to obtain for several reasons. First, one might be interested in a function of causal effects that are available in an existing publication, but the microdata underlying those causal effects might be inaccessible. This could be either because the causal effects are estimated using privately held administrative microdata, as is the case with many of the policies studied by Hendren and Sprung-Keyser (2020), or simply because underlying the replication data wasn't made publicly available. For instance, replication data is not publicly available for nearly half of all empirical papers published in the American Economic Review in recent years (Christensen and Miguel, 2018). Similarly, out of all the papers published in the American Economic Journal: Economic Policy between 2015-2018, approximately 1 out of every 3 papers used either confidential microdata or did not make the replication data public (Herbert, Kingi, Stanchi, and Vilhuber, 2021). In the absence of microdata, it would be infeasible to estimate the correlation across any two causal effects. Second, even if the microdata were available, it might be prohibitively costly to compute the correlation across causal effects when the causal effects rely on different data sources that are challenging to merge together but have common units. This might happen in cases where a unique identifier is missing to exactly merge the two data sources (e.g., in historical decennial Census data) or when the two data sources are housed at different federal agencies (e.g., administrative tax data and administrative crime records for the full population). In these cases, it might be prohibitively difficult to estimate the full correlation structure across causal effects.

In this paper, we provide a simple and easily implementable method to conduct inference on functions of causal effects in the absence of any estimates of the correlation across causal effects. We begin by asking, what is the largest possible variance of the function given the information we observe and what is the correlation structure under which this upper bound is attained? The implied confidence intervals using the upper bound of the variance are conservative but have close to exact coverage when the actual correlation structure is close to the correlation structure under which this upper bound is attained. We show that using this upper bound alone, meaningful inference is possible in settings of interest to applied researchers where estimating the correlation across causal effects is not possible.

Second, we show how inference can be sharpened further when the causal effects being considered correspond to the effect of a randomized treatment on a range of different outcomes. We show that, in this setting, the correlation across the causal effects takes a particularly interpretable form, the sign of which might be known from prior studies, economic theory, or other data sources. Incorporating this information allows us to provide tighter bounds on the variance. Because we cast the problem of finding the upper bound of the variance as an optimization problem, our approach can flexibly incorporate additional setting-specific information to sharpen the variance upper bound, such as the bounded support of an outcome or known independence of two causal effects.

Finally, we re-cast the inference problem as a "breakdown" problem: instead of asking what is the largest possible variance of the function given the available information, we ask how large the variance can be before a policy-relevant conclusion *breaks down*. In the case of MVPF, one policy-relevant null hypothesis is whether a dollar spent on the policy provides the beneficiaries with less than one dollar of benefits, i.e., $H_0 : MVPF < 1$. Our proposed method asks, how plausible is the largest variance under which one can reject this null hypothesis. We provide an easily interpretable breakdown metric that takes the a value between 0 and 1, where 0 implies that one can reject the null hypothesis under any correlation structure and 1 implies that one can't reject the null hypothesis under any correlation structure. A risk-averse policy maker choosing from a menu of policies with equal MVPFs would strictly prefer a policy with a lower breakdown metric.

We illustrate our inference procedure by conducting inference on the Marginal Value of Public Fund for 8 different policies. We show that meaningful inference is possible in the absence of *any* microdata, using the upper bound on the variance alone. For instance, we are able to conclude that increasing expenditure on a universal basic income policy by $1 delivers between $0.77-1.07 of benefits to its recipients. Hendren and Sprung-Keyser (2020) note that since the MVPF reflects the shadow price of redistribution, a welfare-maximizing government should have a positive willingness-to-pay to reduce the statistical uncertainty in the cost of redistribution. We illustrate how statistical uncertainty can be reduced by using setting-specific information about the sign of the correlation across outcomes. Finally, we compute the breakdown metric for the MPVF of multiple policies and show how this metric can be useful to a policymaker choosing from a menu of policies.

The approach of bounding the variance of a scalar-valued parameter in the absence of correlation estimates was recently adopted by Cocci and Plagborg-Møller (2024) in the setting of calibrating parameters of a structural model to match empirical moments. Our work builds upon theirs in three ways. First, while Cocci and Plagborg-Møller (2024) are able to use moment selection tools to tighten the upper bound on the variance, we explicitly characterize the off-diagonal entries of the covariance matrix and show that, in many settings of interest, the sign of the off-diagonal entries will be known from prior studies, economic theory, or other data sources. Second, we characterize the problem of inference under an unknown correlation structure as a breakdown problem (Masten and Poirier, 2020). Since the true correlation structure is unknown, it is unclear how close the upper bound of the variance might be to the true variance. The breakdown approach has the advantage of being comparable across policies. Finally, we illustrate how in settings of interest to applied microeconomists and policymakers, our bounds allow for meaningful inference on functions of causal effects.

2 Setting

Our starting point is a vector of causal effects available in an existing study, $\hat{\beta} \in \mathbb{R}^d$. In our leading application, we are interested in aggregating the vector of causal effects into a scalar estimate of the cost-effectiveness of the policy without access to the microdata using which the estimates $\hat{\beta}$ were constructed. We treat the causal effects to be aggregated as a joint estimate to allow for a distributional theory for a function of causal effects from multiple regression equations.

Assumption 1. *In an existing study, we observe a vector of causal effects* $\hat{\beta} = \left(\hat{\beta}_1, \ldots, \hat{\beta}_d\right)$ *and their corresponding standard errors* $\hat{\sigma} = \left(\hat{\sigma}_1, \ldots, \hat{\sigma}_d\right)$ *such that*

(i) $\hat{\beta} \xrightarrow{p} \beta$

(ii) $\hat{\beta}$ *follows a jointly normal distribution asymptotically,* $\sqrt{n}\left(\hat{\beta} - \beta\right) \xrightarrow{d} \mathcal{N}\left(0, \mathbf{V}\right)$*, where* $\mathbf{V}$ *is a* $d \times d$ *positive semi-definite variance-covariance matrix*

(iii) $\hat{\sigma}$ *is a consistent estimator for the diagonal entries of* $\mathbf{V}$*,* $\left(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_d^2\right) \xrightarrow{p} \left(\sigma_1^2, \ldots, \sigma_d^2\right)$.

Next, we assume that the causal effects $\left(\hat{\beta}_1, \ldots, \hat{\beta}_d\right)$ are aggregated using a possibly non-linear function $f : \mathbb{R}^d \to \mathbb{R}$. This function is assumed to be continuously differentiable at β and $f(\beta)$ is assumed to not be zero-valued. In our application, $f(\hat{\beta})$ corresponds to an estimate of the policy's cost-effectiveness. In our analysis, we place no additional restrictions on the function $f(.)$, since in practice it would be determined using the economics of the specific policy being analyzed.

Assumption 2. *The function* $f : \mathbb{R}^d \to \mathbb{R}$ *is continuously differentiable at* β *and* $f(\beta)$ *is not zero-valued.*

We maintain Assumptions 1 and 2 throughout the paper. Using the delta-method,

$$\sqrt{n}\Big(f(\hat{\beta}) - f(\beta)\Big) \to \mathcal{N}(0, \tau^2)$$

where

$$\tau^2 = \sum_{i=1}^{d}\sum_{j=1}^{d}\sigma_{ij}\frac{\partial f(\beta)}{\partial \beta_i}\frac{\partial f(\beta)}{\partial \beta_j}$$
$$= \sum_{i=1}^{d}\left(\sigma_i\frac{\partial f(\beta)}{\partial \beta_i}\right)^2 + \sum_{\substack{i=1 \\ \{i,j:i\neq j\}}}^{d}\sum_{j=1}^{d}\sigma_{ij}\frac{\partial f(\beta)}{\partial \beta_i}\frac{\partial f(\beta)}{\partial \beta_j} \tag{2.1}$$

and σ_{ij} is the covariance between β_i and β_j.[1]

The objective of this paper is to learn about the variance of $f(\hat{\beta})$. In settings where $f(\hat{\beta})$ is an estimate of the cost-effectiveness of a policy, a policy-maker might also be interested in its variance: a risk-averse policy-maker choosing between implementing two policies that are equally cost-effective would choose to implement the policy with a lower variance of the estimated cost-effectiveness. The central challenge to estimating the variance of $f(\hat{\beta})$ is that it depends on σ_{ij} as can be seen in Equation 2.1, and in many settings of interest, it is not possible to estimate σ_{ij}.

Estimates of σ_{ij} for $i \neq j$ might be difficult to obtain for the following reasons. First, one might be interested in a function of causal effects that are available in an existing publication, but the microdata underlying those causal effects might be inaccessible. This could be either because the causal effects are estimated using privately held administrative microdata or simply because underlying the replication data wasn't made publicly available.[2] In such cases, it would be infeasible to estimate the covariance across any two causal effects. This prevents us from estimating the full variance-covariance matrix, $\mathbf{V}$: while it is standard for papers to report estimates of the diagonal entries (i.e., the standard errors), the off-diagonal entries are rarely reported. Second, even if the microdata were available, it might be prohibitively costly to compute the covariances across causal effects when the causal effects rely on different data sources that are challenging to merge together but have common units. This might happen

[1] The delta-method relies on a linear approximation of the function $f(.)$. However, if the variance is large, the function might be poorly approximated since $\hat{\beta}$ and β would not be close with high probability.

[2] If the microdata underlying all the causal effects were readily accessible, it would be straightforward be able to estimate σ_{ij} for $i \neq j$ (e.g., Zellner, 1962).

in cases where a unique identifier is missing to exactly merge the two data sources (e.g., in historical decennial Census data) or when the two data sources are housed at different federal agencies (e.g., administrative tax data and administrative crime records). In these cases, it might be prohibitively difficult to estimate σ_{ij}. This raises the question, what can be learned about τ^2 when it is not possible to estimate the covariance across any two causal effects? In the next section, we provide an easily implementable inference procedure that allows us to learn about τ^2 in when it is not possible to estimate σ_{ij} for $i \neq j$.

3 Inference Procedure

Since τ^2 depends on σ_{ij} and it is not possible to estimate σ_{ij} for $i \neq j$, we consider bounds on τ^2. Specifically, we consider how large the variance for $f(\hat{\beta})$ can be given the observed information. The motivation for using the upper bound of the variance of $f(\hat{\beta})$ is that any test using this variance guarantees size control and trades it off against a loss of power. Let $\rho_{ij} \equiv \frac{\sigma_{ij}}{\sigma_i \sigma_j}$ be the correlation coefficient between β_i and β_j. We can re-write Equation 2.1 as follows:

$$\tau^2 = \sum_{i=1}^{d} \left(\sigma_i \frac{\partial f(\beta)}{\partial \beta_i} \right)^2 + \sum_{\substack{i=1 \\ \{i,j:i \neq j\}}}^{d} \sum_{j=1}^{d} \rho_{ij} \sigma_i \sigma_j \frac{\partial f(\beta)}{\partial \beta_i} \frac{\partial f(\beta)}{\partial \beta_j} \tag{3.1}$$

Since β_i and σ_i can be consistently estimated given the observed data, obtaining the maximum possible value of τ^2 amounts to maximizing Equation 3.1 with respect to ρ_{ij} for $i \neq j$. We obtain an upper bound on the variance τ^2 using the following optimization problem:

$$\begin{aligned}
&\underset{\{\rho_{ij}\}_{i,j=1}^{d}}{\text{Maximize}} && \tau^2 && && \textbf{SDP.1} \\
&\text{subject to} && \mathbf{V} \succeq 0 && && \text{(C.1)} \\
& && \rho_{ij} = \rho_{ji} && \forall\ i, j = 1, ..., d && \text{(C.2)} \\
& && \rho_{ij} \in [-1, 1] && \forall\ i, j = 1, ..., d && \text{(C.3)}
\end{aligned}$$

C.1 is a matrix inequality that states that the variance-covariance matrix must be positive semi-definite; C.2 requires the resulting variance-covariance matrix to be symmetric; and C.3 ensures that the correlation is bounded between -1 and 1. **SDP.1** is a well-defined semi-definite program (SDP) that can be solved using existing semi-definite programming tools (Vandenberghe and Boyd, 1996).

Absent any additional constraints, the maximum variance is obtained when $\rho_{ij} = 1$ for i, j where $\frac{\partial f(\beta)}{\partial \beta_i}\frac{\partial f(\beta)}{\partial \beta_j} > 0$ and $\rho_{ij} = -1$ if $\frac{\partial f(\beta)}{\partial \beta_i}\frac{\partial f(\beta)}{\partial \beta_j} < 0$. We denote the maximum variance as τ^2_{max}. Then,

$$\begin{aligned} \tau^2_{max} &= \sum_{i=1}^{d} \left(\sigma_i \frac{\partial f(\beta)}{\partial \beta_i} \right)^2 + \sum_{\substack{i=1 \\ \{i,j:i\neq j\}}}^{d} \sum_{j=1}^{d} \sigma_i \sigma_j \left| \frac{\partial f(\beta)}{\partial \beta_i} \frac{\partial f(\beta)}{\partial \beta_j} \right| \\ &= \left(\sum_{i=1}^{d} \sigma_i \left| \frac{\partial f(\beta)}{\partial \beta_i} \frac{\partial f(\beta)}{\partial \beta_j} \right| \right)^2 \end{aligned} \tag{3.2}$$

Some implications arise from inspecting the worst-case variance. First, the confidence intervals constructed using τ_{max} will have a higher coverage probability than when using τ by construction. While this will result in a loss of power, it guarantees size control; the coverage will only be exact when $\tau_{max} = \tau$. Second, it implies a bias-variance tradeoff for measuring the cost-effectiveness of a policy: while adding additional estimates might provide a more accurate view of the cost-effectiveness of a policy, it increases the variance of the cost-effectiveness measure. An implication for applied practice is to measure outcomes that are more closely aligned with the final goal of the policy evaluation. If the goal is evaluate the cost-effectiveness of the policy, measuring outcomes that more closely proxy the cost and benefit of the policy would reduce the variance of the cost-effectiveness measure. Third, even in settings where it is feasible to estimate the covariance across estimates, we recommend that the researcher begin by testing their hypothesis of interest using the upper bound of the variance we provide. Rejecting a null hypothesis using the variance upper bound implies that the hypothesis would also be rejected using the true variance. This allows the researcher to test the hypothesis of interest while also sidestepping the costs of computing the covariances we highlighted above. Finally, the expression for τ_{max} in Equation 3.2 aligns with Lemma 1 in Cocci and Plagborg-Møller (2024). However, given the difference in settings and information available, our work departs from theirs in how we are able to tighten the upper bound on the variance.

3.1 Aggregating Causal Effects from a Randomized Trial

In this sub-section, we specialize to the case where the causal effects being considered correspond to the effect of the same treatment on a range of different outcomes, $Y_1, \ldots, Y_d$. We show that in this setting the off-diagonal entries of $\mathbf{V}$ take a particularly interpretable form, the sign of which might be known from prior work, economic theory, or other data sources. These sign restrictions can be easily incorporated into **SDP.1** to tighten the upper bound on τ^2.

Define $Y_{ij}(1)$ as the treated potential outcome j of unit i and $Y_{ij}(0)$ as the control potential outcome j of unit i, where $j = 1, ..., d$. Let $Z_i \in \{0, 1\}$ denote whether unit i is assigned the treatment or control group. We assume that treatment assignment is independent of the vector of potential outcomes, i.e. $\Big(Y_{ij}(1), Y_{ij}(0)\Big) \perp Z_i$ for all $j = 1, \ldots, d$. The effect of the treatment Z_i is evaluated using d linear regressions:

$$
\begin{aligned}
Y_{i1} &= \alpha_1 + \beta_1 Z_i + \varepsilon_{i1} \\
&\vdots \\
Y_{id} &= \alpha_d + \beta_d Z_i + \varepsilon_{id}
\end{aligned}
$$

Let the vector of Average Treatment Effects (ATEs) be given by $\beta \in \mathbb{R}^d$ and our estimates be given by $\hat{\beta} \in \mathbb{R}^d$. Under our maintained assumptions, we have that

$$
\sqrt{n}(\hat{\beta} - \beta) \to \mathcal{N}(0, \mathbf{V})
$$

Proposition 1. *Under certain regularity conditions, the covariance between the ATE of a* *binary treatment Z_i on outcomes Y_p an Y_q can be expressed as follows:*

$$\mathbb{C}\text{ov}\Big(\beta_p, \beta_q\Big) = \frac{\mathbb{C}\text{ov}(Y_{ip}, Y_{iq} \mid Z_i = 1)}{\mathbb{P}[Z_i = 1]} + \frac{\mathbb{C}\text{ov}(Y_{ip}, Y_{iq} \mid Z_i = 0)}{\mathbb{P}[Z_i = 0]}$$

The proof for Proposition 1 is provided in Appendix Section A. Proposition 1 establishes that covariance of the two ATEs β_j and β_k takes the following interpretable form: if the covariance of the two outcomes is positive in the treatment group *and* the control group, the covariance of the causal effects will also be positive. Intuitively, since the outcomes are positively correlated in treatment and control group, the effect of the treatment must also move in the same direction. Since the direction (and in some cases, the magnitude) of the covariance in the outcomes might be known, this information can easily be added as constraints to **SDP.1** to find a (weakly) smaller upper bound on the variance. For instance, if *all* off-diagonal entries are known to be non-negative, we can solve the following optimization problem:

$$\begin{aligned}
&\underset{\{\rho_{ij}\}_{i,j=1}^{d}}{\text{Maximize}} && \tau^2 && && \textbf{SDP.2} \\
&\text{subject to} && \mathbf{V} \succeq 0 && && \text{(C.1)} \\
& && \rho_{ij} = \rho_{ji} && \forall\, i,j = 1,...,d && \text{(C.2)} \\
& && \rho_{ij} \in [-1, 1] && \forall\, i,j = 1,...,d && \text{(C.3)} \\
& && \rho_{ij} \geq 0 && \forall\, i,j = 1,...,d && \text{(C.4)}
\end{aligned}$$

In Section 5, we show how incorporating sign constraints in real-world applications of interest can yield meaningfully tighter bounds on the variance in some cases.

In some settings, β_i and β_j will identify the ATE of a treatment on outcomes Y_i and Y_j with bounded support. When the outcome has bounded support, the support of the ATE will also bounded. In this case, one might be able to tighten the bounds on the variance further by incorporating this information into the optimization problem **SDP.2**. Hössjer and Sjölander (2022) derive the following bounds for the covariance of two bounded random variables.

Remark 1. *Suppose that outcome* $Y_i \in \left[\underline{Y_i}, \overline{Y_i}\right]$ *and* $Y_j \in \left[\underline{Y_j}, \overline{Y_j}\right]$. *Define* $\underline{\beta_i} = \underline{Y_i} - \overline{Y_i}$ *and* $\overline{\beta_i} = \overline{Y_i} - \underline{Y_i}$. *Define* $\underline{\beta_j}$ *and* $\overline{\beta_j}$ *analogously. Then, the ATEs* β_i *and* β_j *will be bounded:* $\beta_i \in \left[\underline{\beta_i}, \overline{\beta_i}\right]$ *and* $\beta_j \in \left[\underline{\beta_j}, \overline{\beta_j}\right]$. *Then, the covariance of* β_i *and* β_j, σ_{ij} *satisfies:*

$$
\begin{aligned}
&-\min\left[\left(\beta_i - \underline{\beta_i}\right)\left(\beta_j - \underline{\beta_j}\right), \left(\overline{\beta_i} - \beta_i\right)\left(\overline{\beta_j} - \beta_j\right)\right] \\
&\leq \sigma_{ij} \\
&\leq \min\left[\left(\beta_i - \underline{\beta_i}\right)\left(\overline{\beta_j} - \beta_j\right), \left(\overline{\beta_i} - \beta_i\right)\left(\beta_j - \underline{\beta_j}\right)\right].
\end{aligned}
$$

Suppose that the outcomes Y_p, Y_q for $p, q \in \{1, ..., d\}$ are bounded and the bounds are known. Then, we can incorporate the implied bounds on the correlation of β_p and β_q, ρ_{pq}, shown in Remark 1 as a constraint in **SDP.2**. Finally, in some cases, it might be known that two estimates are uncorrelated if, for instance, the estimates are constructed using non-overlapping samples. This information can also incorporated in the optimization problem by fixing that correlation to be 0 as a constraint.

4 Breakdown Analysis

In Section 3, we were concerned with developing an inference procedure for $f(\hat{\beta})$ without knowledge of the full variance-covariance matrix, $\mathbf{V}$. The procedure relies on finding an upper bound of the variance τ^2 to conduct hypothesis tests. In this section, we re-cast the inference problem as a "breakdown" problem: instead of asking what is the largest possible variance of $f(\hat{\beta})$ given the available information, we shift our target to finding the correlation structure across causal effects that yields the largest variance for $f(\hat{\beta})$ *without a conclusion of interest breaking down*. Specifically, we assess the plausibility of the correlation structure across causal effects that maximizes τ^2 *and* under which we can still reject a null hypothesis of interest. In the case of the MVPF, our running example, one policy-relevant null hypothesis is whether a dollar spent on the policy provides the beneficiaries with less than one dollar of benefits, i.e., $H_0 : MVPF < 1$. We provide a breakdown statistic to assesses how "plausible" is the largest variance under which one can reject this null hypothesis – if is deemed to be plausible, the conclusion that the policy is cost-effective would hold under most correlation structures. We formalize our approach in this section.[3]

Let $\boldsymbol{\rho}_{max}$ denote the correlation matrix that maximizes τ^2 in **SDP.1**. Without loss of generality, suppose that we are interested in the hypothesis, $H_0 : f(\hat{\beta}) < k$ vs. $H_1 : f(\hat{\beta}) \geq k$. We break down our approach into the following two steps.

Step 1: Find the correlation structure that maximizes τ^2 *and* under which the 95% CI excludes all values less than k. We will call this correlation matrix $\boldsymbol{\rho}_B$. Specifically, we find $\boldsymbol{\rho}_B$ by solving the following optimization problem:

$$\begin{aligned}
&\underset{\{\rho_{ij}\}_{i,j=1}^{d}}{\text{Maximize}} && \tau^2 && && \textbf{SDP.3}\\
&\text{subject to} && \mathbf{V} \succeq 0 && && \text{(C.1)}\\
& && \rho_{ij} = \rho_{ji} && \forall\ i,j = 1,...,d && \text{(C.2)}\\
& && \rho_{ij} \in [-1, 1] && \forall\ i,j = 1,...,d && \text{(C.3)}\\
& && f(\hat{\beta}) - z_\alpha \times \tau \geq k && && \text{(C.4)}
\end{aligned}$$

where z_α is the $(1 - \alpha)$ quantile of Normal distribution.

Step 2: Assess the plausibility of the correlation matrix, $\boldsymbol{\rho}_B$. To assess the plausibility of this correlation matrix, we calculate the following statistic:

$$\text{Breakdown Statistic} = \begin{cases} \dfrac{\|\boldsymbol{\rho}_B - \boldsymbol{\rho}_{max}\|_F}{2\sqrt{d(d-1)}} & \text{if there exists a feasible solution to } \textbf{SDP.3} \\ 1 & \text{if there is no feasible solution to } \textbf{SDP.3} \end{cases}$$

where $\|.\|_F$ is the Frobenius norm.

[3]This approach of assessing the plausibility of the point at which a conclusion of interest "breaks down" is referred to as breakdown analysis. See Masten and Poirier (2020) and Spini (2021) for recent examples.

We refer to this number as the Breakdown Statistic. Note that it is bounded below by 0 and 1. A breakdown statistic of 0 implies that we can reject the null hypothesis under any correlation matrix. The breakdown statistic is 0 when $\boldsymbol{\rho}_B = \boldsymbol{\rho}_{max}$. This implies that, even under the worst case correlation structure $\boldsymbol{\rho}_{max}$, we can reject the null hypothesis that $H_0 : f(\hat{\beta}) < k$. In other words, there is no correlation structure under which we can't reject this null hypothesis. A breakdown statistic of 0.5 implies that the policy conclusion is valid under independence of all effects. The breakdown statistic takes the value 0.5 when $\boldsymbol{\rho}_B = I_d$, i.e., when all estimates are uncorrelated. If there is no feasible solution to **SDP.3**, we set the breakdown statistic to 1. A breakdown statistic of 1 implies that the policy conclusion is not valid under any correlation structure.

5 Application

In this section, we illustrate how the methods we develop in Sections 3 and 4 can be used to achieve meaningful inference in settings of interest to economists and policymakers alike without access to any microdata and without any knowledge of the off-diagonal entries of $\mathbf{V}$. We apply our inference method to the Marginal Value of Public Funds (MVPF) framework. First, we briefly describe the MVPF framework. Second, we argue why our approach to inference is particularly well-suited for the MVPF framework. Third, we illustrate our method in the context of 8 different public policies.

Hendren and Sprung-Keyser (2020) popularized the MVPF as a unified metric for evaluating the welfare consequences of government policy. For any policy change, the MVPF is the ratio of the benefits that the policy provides to its recipients, divided by the policy's net cost to the government. Specifically,

$$MVPF = \frac{Benefits}{Net\ Government\ Costs} = \frac{\Delta W}{\Delta E - \Delta C}$$

where ΔW denotes the benefits that the policy provides to individuals in the population, ΔE denotes the government's upfront expenditure on the policy, and ΔC denotes the long-run reduction in government costs due to the the causal effect of the policy. The MVPF measures a policy's "bang for the buck". For a given policy with $MVPF = A$, the policy delivers \$A of benefits per dollar of net government spending.

Four features of the MVPF framework make our inference methods particularly desirable. First, the MVPF for a given policy is constructed as a non-linear function of multiple causal effects. Consider the MVPF of an expanded Earning Income Tax Credit (EITC) program, Paycheck Plus, which provides tax credits to low-income singles who traditionally are not eligible for significant EITC benefits. Miller, Katz, Azurdia, Isen, and Schultz (2017) calculate the causal effect of the tax credit on earnings, employment, and after-tax income. The estimates used to construct the MVPF the only information available for inference is summarized in Table 1. Adapting to our notation,

$$\begin{aligned}\hat{\beta} &= \begin{bmatrix}\hat{\beta}_1 & \hat{\beta}_2 & \hat{\beta}_3 & \hat{\beta}_4 & \hat{\beta}_5 & \hat{\beta}_6\end{bmatrix}' \\ &= \begin{bmatrix}0.009 & 0.025 & 654 & 33 & 645 & 192\end{bmatrix}'\end{aligned}$$

and

$$\begin{aligned}MVPF_{\text{Paycheck Plus}} \equiv f(\hat{\beta}) &= \frac{1399 \times (45 - \hat{\beta}_1) + 1364 \times (34.8 - \hat{\beta}_2)}{(\hat{\beta}_3 - \hat{\beta}_4) + (\hat{\beta}_5 - \hat{\beta}_6)} \\ &= 0.996.\end{aligned}$$

Second, in most cases, the only information available to construct the MVPF are the causal effects and their corresponding standard errors. In the case of Paycheck Plus, the estimates are constructed using administrative tax data for the experimental sample. Since this data is not publicly accessible and the authors of the study do not report the correlation across causal effects, we have access to estimates only for the diagonal entries of the variance-covariance matrix, $\mathbf{V}$. Therefore, the only information we have to conduct inference on the MVPF of Paycheck Plus is what is provided in Table 1.[4]

Third, Hendren and Sprung-Keyser (2020) show that increasing spending on Policy A is welfare-improving by reducing spending on Policy B if and only if the MVPF of Policy A is greater than the MVPF of Policy B. This presents a natural null hypothesis central to welfare analysis and policy choice, $H_0 : MVPF_A < MVPF_B$.[5] To construct such a test that controls size under any feasible correlation structure, one would need to solve the optimization problems provided in Section 3.

Fourth, Hendren and Sprung-Keyser (2020) note that since the MVPF reflects the shadow price of redistribution, a welfare-maximizing government should be willing to pay to reduce the statistical uncertainty in the cost of redistribution. Our approach allows us to first construct the confidence intervals that control size under any feasible correlation structure. In Section 3.1, we show how mild setting-specific assumptions can be used to sharpen inference on the MVPF in many settings, establishing a systematic approach to reducing statistical uncertainty in the MVPF estimates.

We illustrate our inference method through the MVPF of 8 different policies. We consider policies from different domains of public expenditure: three job-training programs (Job Start, Work Advance, Year Up), two cash transfers (Paycheck Plus, Alaska Universal Basic Income), a health insurance expansion (Medicare Part D), childcare expenditure (foster care provision), and an Unemployment Insurance (UI) expansion. The MVPF and the 95% confidence intervals constructed using **SDP.1** are shown in Figure 1. We defer the details of each policy and their corresponding MVPF calculations to the Appendix. Some key lessons emerge from examining Figure 1. First, absent *any* assumptions on the off-diagonal entries of variance-covariance matrix, $\mathbf{V}$, we are able to reject the null hypothesis that MVPF of two job training programs (Job Start and Year Up) is greater than 1 under any correlation

[4] To conduct inference on MVPF estimates, Hendren and Sprung-Keyser (2020) circumvent the problem of an unknown correlation structure by assuming a correlation structure across estimates. However, this approach might lead to confidence intervals that don't have the appropriate coverage probability.

[5] Here, we assume that the beneficiaries of all policies receive equal welfare weights.

structure across causal effects. Specifically, we can conclude that $1 spent on a job training program provides the recipients of the policy with less than $1 of benefits. Second, comparing across policies, we can additionally conclude that the MVPF of Alaska UBI is greater than the MVPF of Year Up, under *any* correlation matrix; this implies that reducing expenditure on Year Up and increasing expenditure on Alaska UBI would be welfare-improving. Third, the only instance in which we have access to the microdata is for the causal effects underlying the MVPF for Medicaid Par D. Using this microdata, we are able to estimate the full variance-covariance matrix, $\mathbf{V}$, and compare confidence intervals with exact coverage that are infeasible to compute in most cases to the conservative confidence intervals resulting from **SDP.1**. While the exact confidence intervals rule out values of the MVPF smaller than 0.80 and larger than 1.95, the conservative confidence intervals only somewhat wider allowing us to rule out values less than 0.17 and larger than 2.57. This suggests that one can conduct meaningful inference on the MVPF for a given policy using the widest possible confidence intervals alone. Finally, for most of the policies we consider, we are unable to reject the null that the policies are not cost-effective, $H_0 : MVPF < 1$. This suggests that there might be meaningful statistical uncertainty in the estimates for the MVPF. Absent estimates of the variance for the MVPF of the policies we consider, a policymaker might conclude that increasing funding for UI extensions by decreasing funding for job training programs will be welfare improving. However, our results suggest that the statistical uncertainty in the estimates doesn't allow us to conclude this from the available information.

Since we are unable to reject the null that the policies are not cost-effective, $H_0 : MVPF < 1$, one might ask how far would we need to move away from the worst case correlation structure to be able to reject this null hypothesis. The Breakdown Statistic proposed in Section 4 measures how far one would need to move from the worst case to be able to reject the null hypothesis – the higher the Breakdown Statistic, the further one would need to move from the worst case correlation structure. Clearly, if the MVPF point estimate is lesser than 1, we cannot reject this null hypothesis under any correlation structure and the Breakdown Statistic is equal to 1. We construct the Breakdown Statistic for the introduction of Medicare Part D, foster care provision, and UI extension, fixing $k = 1$ in Section 4. The Breakdown Statistic for the policies is shown in Table 2. Comparing the breakdown statistic for Medicare Part D (0.60) and Foster Care Provision (0.83), we can conclude that one would have to move farther from the worst case correlation structure in the case of Foster Care Provision to be able to reject the null that the policy delivers less than $1 of benefits per dollar of net government spending.

Next, we restrict our attention to the policies evaluated using randomized trials, the setting covered in Section 3.1: Job Start, Paycheck Plus, Work Advance, and Year Up. In each case, we add additional sign restrictions to the optimization problem, following our result in 1. In the case of Paycheck Plus, for instance, it seems plausible to assume that individuals with higher after-tax income are also more likely to have higher earnings and more likely to participate in the labor force within the treatment group and the control group. Therefore, we compute the confidence intervals for the Paycheck Plus MVPF using **SDP.2**. Figure 2 summarizes the results by including additional sign constraints. As mentioned earlier, a welfare-maximizing government should be willing to pay to reduce the statistical uncertainty in the cost of redistribution. Figure 2 shows that including the additional sign constraints implied by 1 can meaningfully sharpen inference. For instance, in the case of Paycheck Plus, we can now rule out values for the MVPF smaller than -0.38 and larger than 2.37.

Table 1: MVPF Calculation for Paycheck Plus

	(1) Year	(2) Estimate	(3) SE
Average Bonus Paid	2014	1399	
Average Bonus Paid	2015	1364	
Take-Up	2014	45.90%	
Take-Up	2015	34.80%	
Extensive Margin Labor Market ($\hat{\beta}_1$)	2014	0.90%	0.65%
Extensive Margin Labor Market ($\hat{\beta}_2$)	2015	2.5%	0.91%
Impact on After Tax Income ($\hat{\beta}_3$)	2014	654	187.79
Impact on Earnings ($\hat{\beta}_4$)	2014	33	43.35
Impact on After Tax Income ($\hat{\beta}_5$)	2015	645	241.15
Impact on Earnings ($\hat{\beta}_6$)	2015	192	177.71
WTP		1071	
Net Government Costs		1074	
MVPF		0.996	

Table 2: Breakdown Statistics for MVPF

	(1) Breakdown Statistic
Medicare Part D	0.60
Foster Care Provision	0.83
UI Extension	0.62

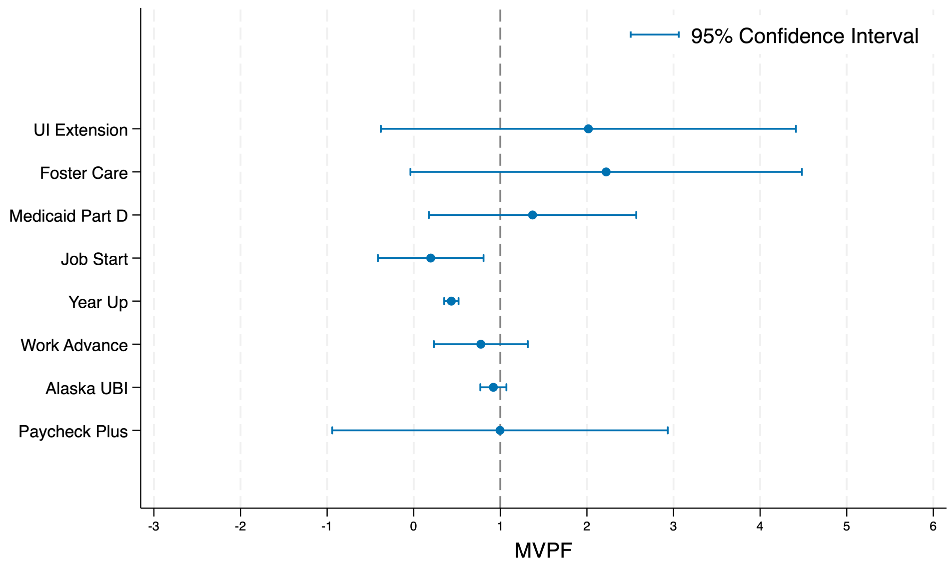

Figure 1: MVPF Confidence Intervals

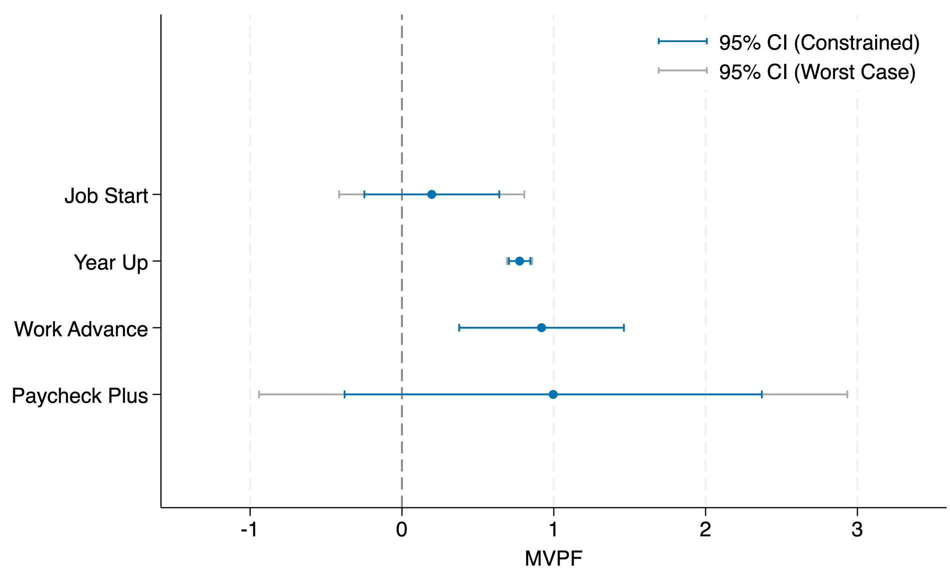

Figure 2: MVPF Confidence Intervals with Sign Constraints

References

Christensen, G., and E. Miguel (2018): "Transparency, reproducibility, and the credibility of economics research," *Journal of Economic Literature*, 56(3), 920–980. 1

Cocci, M. D., and M. Plagborg-Møller (2024): "Standard errors for calibrated parameters," *Review of Economic Studies*. 1, 3

Hansen, B. (2022): *Econometrics*. Princeton University Press. A

Hendren, N., and B. Sprung-Keyser (2020): "A unified welfare analysis of government policies," *The Quarterly Journal of Economics*, 135(3), 1209–1318. 1, 5, 4

Herbert, S., H. Kingi, F. Stanchi, and L. Vilhuber (2021): "The reproducibility of economics research: A case study," *Banque de France Working Paper, Forthcoming*. 1

Hössjer, O., and A. Sjölander (2022): "Sharp lower and upper bounds for the covariance of bounded random variables," *Statistics & Probability Letters*, 182, 109323. 3.1

Masten, M. A., and A. Poirier (2020): "Inference on breakdown frontiers," *Quantitative Economics*, 11(1), 41–111. 1, 3

Miller, C., L. F. Katz, G. Azurdia, A. Isen, and C. B. Schultz (2017): "Expanding the earned income tax credit for workers without dependent children: Interim findings from the paycheck plus demonstration in new york city," *New York: MDRC, September*. 5

Spini, P. E. (2021): "Robustness, heterogeneous treatment effects and covariate shifts," *arXiv preprint arXiv:2112.09259*. 3

Vandenberghe, L., and S. Boyd (1996): "Semidefinite programming," *SIAM review*, 38(1), 49–95. 3

Zellner, A. (1962): "An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias," *Journal of the American statistical Association*, 57(298), 348–368. 2

Appendix

Table of Contents

A Proof of Proposition 1 A.2

A Proof of Proposition 1

Define $Y_{ij}(1)$ as the treated potential outcome j of unit i and $Y_{ij}(0)$ as the control potential outcome j of unit i, where $j = 1, ..., d$. Let $Z_i \in \{0, 1\}$ denote whether unit i is assigned the treatment or control group. We assume that treatment assignment is independent of the vector of potential outcomes, i.e. $\Big(Y_{ij}(1), Y_{ij}(0)\Big) \perp Z_i$ for all $j = 1, \ldots, d$. We decompose the potential outcomes for each j into an expectation component and an error term:

$$
\begin{aligned}
Y_{ij}(1) &= \mathbb{E}[Y_{ij}(1)] + \varepsilon_{ij}(1) \\
Y_{ij}(0) &= \mathbb{E}[Y_{ij}(0)] + \varepsilon_{ij}(0)
\end{aligned}
$$

where $\varepsilon_{ij}(1) \equiv Y_{ij}(1) - \mathbb{E}[Y_{ij}(1)]$ and $\varepsilon_{ij}(0) \equiv Y_{ij}(0) - \mathbb{E}[Y_{ij}(0)]$. We rewriting the observed outcome variable, Y_{ij}, as follows.

$$
\begin{aligned}
Y_{ij} &= Z_i Y_{ij}(1) + (1 - Z_i)\, Y_{ij}(0) \\
&= Z_i \Big(\mathbb{E}\left[Y_{ij}(1)\right] + \varepsilon_{ij}(1)\Big) + \Big(1 - Z_i\Big)\Big(\mathbb{E}\left[Y_{ij}(0)\right] + \varepsilon_{ij}(0)\Big) \\
&= \mathbb{E}\left[Y_{ij}(0)\right] + Z_i \Big(\mathbb{E}\left[Y_{ij}(1)\right] - \mathbb{E}\left[Y_{ij}(0)\right]\Big) + Z_i \varepsilon_{ij}(1) + (1 - Z_i)\, \varepsilon_{ij}(0) \\
&= \alpha_j + \beta_j Z_i + \varepsilon_{ij}
\end{aligned}
$$

where $\alpha_j \equiv \mathbb{E}\left[Y_{ij}(0)\right], \beta_j \equiv \mathbb{E}\left[Y_{ij}(1) - Y_{ij}(0)\right]$, and $\varepsilon_{ij} \equiv Z_i \varepsilon_{ij}(1) + (1 - Z_i)\, \varepsilon_{ij}(0)$. Suppose that effect of the treatment Z_i is evaluated using d linear regressions:

$$
\begin{aligned}
Y_{i1} &= \alpha_1 + \beta_1 Z_i + \varepsilon_{i1} \\
&\vdots \\
Y_{id} &= \alpha_d + \beta_d Z_i + \varepsilon_{id}
\end{aligned}
$$

Let the vector of Average Treatment Effects (ATEs) be given by $\beta \in \mathbb{R}^d$ and our estimates be given by $\hat{\beta} \in \mathbb{R}^d$. Under our maintained assumptions, we have that

$$\sqrt{n}(\hat{\beta} - \beta) \to \mathcal{N}(0, \mathbf{V})$$

For any $p \neq q$, the (p,q)-th entry of the variance-covariance matrix $\mathbf{V}$ can be expressed as follows (Hansen, 2022):

$$\mathbb{C}\text{ov}[\beta_p, \beta_q] = \frac{\mathbb{E}[\varepsilon_{ip}\varepsilon_{iq}]\mathbb{E}[Z_i]^2 - \mathbb{E}[\varepsilon_{ip}\varepsilon_{iq}Z_i]\mathbb{E}[Z_i] - \mathbb{E}[\varepsilon_{ip}\varepsilon_{iq}Z_i]\mathbb{E}[Z_i] + \mathbb{E}[\varepsilon_{ip}\varepsilon_{iq}Z_i]}{\mathbb{E}[Z_i]^2(1 - \mathbb{E}[Z_i])^2} \tag{A.1}$$

Note that the term $\mathbb{E}[\varepsilon_{ip}\varepsilon_{iq}Z_i]$ can be simplified as follows:

$$\begin{aligned}
\mathbb{E}[\varepsilon_{ip}\varepsilon_{iq}Z_i] &= \mathbb{E}[\varepsilon_{ip}\varepsilon_{iq} \mid Z_i = 1]\mathbb{E}[Z_i] \\
&= \mathbb{E}\Big[Y_{ip}(1)Y_{iq}(1) - Y_{ip}(1)\mathbb{E}[Y_{iq}(1)] - \mathbb{E}[Y_{ip}(1)]Y_{iq}(1) + \mathbb{E}[Y_{ip}(1)]\mathbb{E}[Y_{iq}(1)] \mid Z_i = 1\Big]\mathbb{E}[Z_i] \\
&= \Big(\mathbb{E}[Y_{ip}(1)Y_{iq}(1) \mid Z_i = 1] - \mathbb{E}[Y_{ip}(1) \mid Z_i = 1]\mathbb{E}[Y_{iq}(1) \mid Z_i = 1]\Big)\mathbb{E}[Z_i] \\
&= \mathbb{E}[Z_i]\mathbb{C}\text{ov}\Big(Y_{ip}(1), Y_{iq}(1) \mid Z_i = 1\Big) = \mathbb{E}[Z_i]\mathbb{C}\text{ov}\Big(Y_{ip}, Y_{iq} \mid Z_i = 1\Big)
\end{aligned}$$

where the first equality follows from the Law of Total Probability and the third equality follows from random assignment of the treatment Z_i. Similarly, $\mathbb{E}[\varepsilon_{ip}\varepsilon_{iq}]$ be simplified as follows:

$$
\begin{aligned}
\mathbb{E}[\varepsilon_{ip}\varepsilon_{iq}] &= \mathbb{E}\left[\Big(Z_i\varepsilon_{ip}(1) + (1-Z_i)\,\varepsilon_{ip}(0)\Big)\Big(Z_i\varepsilon_{iq}(1) + (1-Z_i)\,\varepsilon_{iq}(0)\Big)\right] \\
&= \mathbb{E}\Big[Z_i\varepsilon_{ip}(1)\varepsilon_{iq}(1)\Big] + \mathbb{E}\Big[\,(1-Z_i)\,\varepsilon_{ip}(0)\varepsilon_{iq}(0)\Big] \\
&= \mathbb{E}\Big[Z_i\Big(Y_{ip}(1) - \mathbb{E}[Y_{ip}(1)]\Big)\Big(Y_{iq}(1) - \mathbb{E}[Y_{iq}(1)]\Big)\Big] \\
&\qquad + \mathbb{E}\Big[\,(1-Z_i)\,\Big(Y_{ip}(0) - \mathbb{E}[Y_{ip}(0)]\Big)\Big(Y_{iq}(0) - \mathbb{E}[Y_{iq}(0)]\Big)\Big] \\
&= \mathbb{E}\Big[\Big(Y_{ip}(1) - \mathbb{E}[Y_{ip}(1)]\Big)\Big(Y_{iq}(1) - \mathbb{E}[Y_{iq}(1)]\Big) \mid Z_i = 1\Big]\mathbb{E}[Z_i] \\
&\qquad + \mathbb{E}\Big[\Big(Y_{ip}(0) - \mathbb{E}[Y_{ip}(0)]\Big)\Big(Y_{iq}(0) - \mathbb{E}[Y_{iq}(0)]\Big) \mid Z_i = 0\Big]\mathbb{E}[1-Z_i] \\
&= \mathbb{E}[Z_i]\mathbb{C}\text{ov}\Big(Y_{ip}(1), Y_{iq}(1) \mid Z_i = 1\Big) + \Big(1 - \mathbb{E}[Z_i]\Big)\mathbb{C}\text{ov}\Big(Y_{ip}(0), Y_{iq}(0) \mid Z_i = 0\Big) \\
&= \mathbb{E}[Z_i]\mathbb{C}\text{ov}\Big(Y_{ip}, Y_{iq} \mid Z_i = 1\Big) + \Big(1 - \mathbb{E}[Z_i]\Big)\mathbb{C}\text{ov}\Big(Y_{ip}, Y_{iq} \mid Z_i = 0\Big)
\end{aligned}
$$

Plugging the above into A.1 and simplifying we obtain,

$$
\mathbb{C}\text{ov}[\beta_p, \beta_q] = \frac{\mathbb{C}\text{ov}(Y_{ip}, Y_{iq} \mid Z_i = 1)}{\mathbb{E}[Z_i]} + \frac{\mathbb{C}\text{ov}(Y_{ip}, Y_{iq} \mid Z_i = 0)}{1 - \mathbb{E}[Z_i]}
$$