Inference for Welfare Metrics*

Vedant Vohra

February 17, 2026

Abstract

Economists often estimate causal effects of policies on multiple outcomes and summarize them into scalar measures of cost-effectiveness or welfare, such as the Marginal Value of Public Funds (MVPF). In many settings, microdata underlying these estimates are unavailable, leaving researchers with only published estimates and their standard errors. We develop tools for valid inference on functions of causal effects, such as the MVPF, when the correlation structure is unknown. Our approach is to construct worst-case confidence intervals, leveraging experimental designs to tighten them, and to assess robustness using breakdown analyses. We illustrate our method with MVPFs for eight policies. (*JEL* C12, C21, H00)

*Vohra (vohra@ucsd.edu): UC San Diego. I am grateful to Julie Cullen, Graham Elliott, Itzik Fadlon, Stefan Faridani, Jacob Goldin, Xinwei Ma, Craig McIntosh, Harrison Mitchell, and Kaspar Wuthrich for helpful comments. I also thank the editor, Danny Yagan, and three anonymous referees for valuable comments and suggestions. Earlier drafts of this paper were titled "Valid Inference on Functions of Causal Effects in the Absence of Microdata" and "Inference for the Marginal Value of Public Funds."

1 Introduction

Policymakers increasingly rely on evidence from randomized evaluations to guide decisions about which programs to fund and scale. These evaluations often report multiple estimated causal effects—for example, a tax credit program's impacts on earnings, after-tax income, and labor force participation—but policymakers typically care about a summary measure that aggregates these effects into a measure of overall cost-effectiveness of the policy, such as the Marginal Value of Public Funds (MVPF). Determining whether a program delivers "bang for the buck" thus requires inference on a scalar function of several estimated causal effects, rather than on any individual effect alone.

Conducting valid inference on such summary measures of policy cost-effectiveness is often difficult in practice. In many cases, researchers observe only the estimated causal effects and their standard errors but lack information about the correlations between them. These correlations are crucial for quantifying uncertainty: the variance of any scalar function that aggregates multiple effects depends not only on the precision of each individual estimate but also on how the estimates co-vary. If the underlying microdata were available, the covariances between causal effect estimates could be computed directly, allowing the variance of the function to be estimated (e.g., Zellner, 1962). However, in many settings of practical interest, the microdata are unavailable for ex-post analysis, leaving researchers to rely only on published estimates and their standard errors. In this paper, we study the problem of inference on functions of multiple causal effects when the correlation structure across these causal effects is unknown.

To illustrate the challenge, we focus on the problem of conducting inference for the Marginal Value of Public Funds (MVPF) of a policy (Hendren and Sprung-Keyser, 2020). The MVPF is a widely used metric for evaluating the welfare consequences of government expenditure. It is defined as a non-linear function of multiple causal effects: the benefits a policy provides to its recipients are divided by the policy's net cost to the government. Hendren and Sprung-Keyser (2020) construct MVPFs for more than one hundred policies using causal effects reported in existing studies. In most cases, only the point estimates and their standard errors are available to them, while the microdata underlying these estimates are inaccessible for such ex-post analysis. The challenge for inference is that the variance of the MVPF depends on the correlations across causal effects, which are not reported and not estimable.

We propose a simple inference procedure that delivers valid confidence intervals for functions of causal effects, even when the correlation structure across effects is unknown. The idea is straightforward: we ask what is the largest possible variance of the function given the available information and we identify the correlation structure under which this upper bound is attained. Using this worst-case variance, we construct conservative confidence intervals that guarantee valid coverage. We show that this conservative approach enables meaningful inference when computing the confidence intervals for MVPF estimates. Importantly, we formulate the problem of finding the variance upper bound as an optimization problem, which also makes it straightforward to incorporate other setting-specific information—for example, known independence between causal effects—to further tighten the confidence intervals and improve statistical precision.

Second, we show how confidence intervals can be tightened further still when the causal effects correspond to the impacts of a randomized treatment on multiple outcomes. In this setting, we characterize the off-diagonal entries of the covariance matrix and show that they take a particularly interpretable form, the sign of which may be known from prior studies, economic theory, or other data sources. Incorporating such information allows us to meaningfully increase statistical power to reject null hypotheses of interest.

Finally, we introduce a complementary approach to inference and ask a different question from worst-case inference. Instead of focusing on the largest possible variance given the available information, we ask how sensitive a policy-relevant conclusion is to uncertainty about the correlation structure. For example, a policymaker may wish to test whether a policy "pays for itself." Bergstrom, Dodds and Rios (2025) argue that, among policies with positive willingness-to-pay, those that pay for themselves (i.e., have negative net fiscal costs) should be assigned negative MVPFs. This motivates testing whether the MVPF is strictly negative for such policies. If one rejects the null hypothesis H_0 : MVPF < 0 against H_1 : MVPF ≥ 0, one may conclude that policy does *not* pay for itself. We introduce a "breakdown statistic" that quantifies how robust this conclusion is to different correlation structures: it measures the proportion of admissible correlation structures under which the null hypothesis would *not* be rejected. The statistic takes values between 0 and 1, where a value of 0 implies that we can conclude that the MVPF is greater than 0 under all plausible correlation structures, while a value of 1 implies that we cannot reject the null under any correlation structure. Unlike inference based on the worst-case variance, which guarantees valid coverage but might be conservative, the breakdown statistic facilitates comparisons of the robustness with which we can arrive at a policy conclusion—for example, whether the MVPF of a policy is greater than 0—across settings.

We illustrate our inference procedure by conducting inference on the MVPF for eight different policies. First, we show that meaningful inference is possible even in the absence of *any* microdata, using the upper bound of the variance alone. Second, Hendren and Sprung-Keyser (2020) note that because the MVPF reflects the shadow price of redistribution, a welfare-maximizing government should have a positive willingness-to-pay to reduce the statistical uncertainty around the cost of redistribution. We demonstrate how this uncertainty can be reduced by leveraging setting-specific information about the sign of correlations across outcomes. In fact, our novel characterization of the covariance structure in randomized trials allows us to tighten MVPF confidence intervals beyond the worst-case by up to 30% in the policies we consider. Finally, we compute the breakdown statistic for the MVPF across multiple policies and illustrate how this metric can guide policymakers choosing among alternative policies.

Our work contributes to the literature on welfare analyses of government expenditure (e.g., Chetty, 2009; Heckman et al., 2010; Hendren and Sprung-Keyser, 2020). While existing tools provide a unified framework for evaluating the welfare consequences of government policies, statistical methods for conducting inference on welfare metrics under frequently encountered data limitations have been less developed. Our inference procedures strengthen the MVPF framework by providing a formal approach to quantifying statistical uncertainty in welfare metrics. Hendren and Sprung-Keyser (2020) show that, when social welfare weights are equal across beneficiaries of two policies, a marginal reallocation of spending from Policy B to Policy A is welfare-improving if and only if the MVPF of Policy A exceeds that of Policy B. The methods developed in this paper therefore provide a valid test of this policy-relevant null hypothesis, $H_0 : \text{MVPF}_A \leq \text{MVPF}_B$.

Hendren and Sprung-Keyser (2020) propose a parametric bootstrap approach that constructs confidence intervals for the MVPF under a user-specified correlation structure. While such an approach can yield valid inference when the specified structure is indeed the worst case, misspecification may lead to confidence intervals that fail to achieve nominal coverage. Our method avoids this risk by formally identifying—rather than assuming—the correlation structure that maximizes the variance, solving an optimization problem that guarantees valid inference regardless of the true correlation structure. Moreover, our setup allows us to incorporate additional setting-specific information—for example, known independence across estimates or theory-driven sign restrictions—thereby increasing statistical power when such information is available.

At the same time, we caution that worst-case confidence intervals correspond to the most adversarial correlation structure consistent with the imposed constraints. In many settings, the worst-case confidence intervals implied by our method may be overly conservative—even after imposing theory-informed sign restrictions—relative to the true underlying uncertainty faced by applied researchers. A central takeaway for practitioners, therefore, is that estimating and reporting the covariance matrix across causal effects, when feasible, can substantially sharpen ex-post inference for welfare metrics constructed from those effects. If consistently estimating the covariance matrix is infeasible but additional information is available—for example, through auxiliary data or features of the research design that can be used to estimate or bound correlations across outcomes—researchers can incorporate these restrictions into the optimization problem, making inference less conservative.

Our methods might be applicable beyond the MVPF as well, in other settings when the correlation structure across causal effects might be difficult to obtain. First, researchers are frequently interested in functions of causal effects reported in existing publications, but the underlying microdata may be inaccessible. This can occur when the effects are estimated using privately held administrative data or when replication files are not publicly released. Replication data are missing for nearly half of all empirical papers published in the *American Economic Review* (Christensen and Miguel, 2018), underscoring how common this problem is. Second, even when the underlying data are technically available, computing the correlations can be prohibitively costly when the effects come from distinct datasets with common units but difficult-to-merge identifiers. For example, unique identifiers may be missing in historical decennial Census data (Ruggles, Fitch and Roberts, 2018), or the relevant data sources may be stored across separate federal agencies, as is the case when linking administrative tax data and administrative crime records for the full population (Rose, 2018).

Finally, Cocci and Plagborg-Møller (2024) develop a closely related procedure in a different context: bounding the asymptotic variance of an estimate for a structural parameter when calibrating models to match empirical moments. Their work shows how to compute worst-case standard errors when the off-diagonal elements of the variance-covariance matrix are unknown, using only the variances of the empirical moments. We extend their framework in three important ways. First, we exploit the structure of randomized treatments to characterize the covariance matrix, which enables us to impose theory-motivated sign restrictions and potentially yields substantial power improvements. Second, our paper applies this variance-bounding approach to a new domain—inference on welfare metrics, such as the MVPF—where analysts frequently lack access to the underlying microdata. Third, we introduce a "breakdown" approach that quantifies the robustness of policy conclusions to uncertainty about the correlation structure.

2 Setting

Our starting point is a vector of estimated causal effects, denoted by $\hat{\boldsymbol{\beta}} \in \mathbb{R}^d$, that we seek to aggregate into a measure of the cost-effectiveness of a policy. We assume that $\hat{\boldsymbol{\beta}}$ asymptotically follows a joint Normal distribution with variance-covariance matrix $\mathbf{V}$. Since it is standard practice to report standard errors for individual estimates, we assume we have access to consistent estimates of the diagonal entries of $\mathbf{V}$. In contrast, the covariances between estimated causal effects—the off-diagonal entries of $\mathbf{V}$—are rarely reported. We focus on a setting where the underlying microdata are unavailable, so these off-diagonal entries cannot be directly estimated. We summarize the available information in Assumption 1.

Assumption 1. *From an existing study, we observe a vector of estimated causal effects* $\hat{\boldsymbol{\beta}} = \big(\widehat{\beta}_1, \ldots, \widehat{\beta}_d\big)$ *for the true causal effects* $\boldsymbol{\beta} = \big(\beta_1, \ldots, \beta_d\big)$*, along with their corresponding standard errors* $\hat{\boldsymbol{\sigma}} = \big(\widehat{\sigma}_1, \ldots, \widehat{\sigma}_d\big)$*. We assume:*

(i) Consistency: $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}$.

(ii) Asymptotic normality: $\sqrt{n}\big(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\big) \xrightarrow{d} \mathcal{N}\big(0, \mathbf{V}\big)$*, where* $\mathbf{V}$ *is a* $d \times d$ *positive semi-definite variance-covariance matrix.*

(iii) Consistent standard errors: $\hat{\boldsymbol{\sigma}}^{\mathbf{2}}$ *consistently estimates the diagonal entries of* $\mathbf{V}$*, i.e.* $\big(\widehat{\sigma}_1^2, \ldots, \widehat{\sigma}_d^2\big) \xrightarrow{p} \big(\sigma_1^2, \ldots, \sigma_d^2\big)$.

In our setting, $f(\hat{\boldsymbol{\beta}})$ represents an estimate of a policy's cost-effectiveness. We assume that the function aggregating the causal effects, $f : \mathbb{R}^d \to \mathbb{R}$, is continuously differentiable at $\boldsymbol{\beta}$, with $f'(\boldsymbol{\beta}) \neq 0$. We place no additional restrictions on $f(\cdot)$ and allow it to be non-linear, since in practice its form depends on the economic mapping between the estimated causal effects and the cost-effectiveness measure for the policy being analyzed. This is summarized in Assumption 2.

Assumption 2. *The function* $f : \mathbb{R}^d \to \mathbb{R}$ *is continuously differentiable at* $\boldsymbol{\beta}$ *and satisfies* $f'(\boldsymbol{\beta}) \neq 0$.

In summary, Assumption 1 states that we observe consistent estimates of the causal effects and their standard errors, but lack reliable information about the covariances between them. Assumption 2 requires that the function mapping estimated causal effects into the cost-effectiveness measure is smooth and well-behaved. We maintain Assumptions 1 and 2 throughout the paper.

Under these assumptions, we apply the delta method to obtain the asymptotic distribution of $f(\hat{\boldsymbol{\beta}})$:

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

where

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

and σ_{ij} denotes the covariance between β_i and β_j.[1] We define

$$\rho_{ij} \equiv \frac{\sigma_{ij}}{\sigma_i\sigma_j},$$

as the correlation coefficient between β_i and β_j.

The objective of this paper is to learn about τ^2, the asymptotic variance of $f(\hat{\boldsymbol{\beta}})$. The central challenge is that τ^2 depends on the covariances σ_{ij}, which are not estimable in our setting because the underlying microdata are unavailable.

3 Inference Procedure

Since the asymptotic variance of $f(\hat{\boldsymbol{\beta}})$ depends on the correlation structure across the estimated causal effects—and this correlation structure cannot be estimated in the absence of

[1]The delta method relies on a first-order (linear) approximation of the function $f(\cdot)$ around $\boldsymbol{\beta}$. Under the maintained assumptions, this approximation is valid asymptotically. However, if the variance of $\hat{\boldsymbol{\beta}}$ is large, higher-order terms in the Taylor expansion may be non-negligible making the linear approximation less accurate (see, e.g., Section 3.3 in van der Vaart, 1998).

microdata—we consider an alternative approach to inference on $f(\boldsymbol{\beta})$. We ask: given the observed information, how large could the asymptotic variance of $f(\hat{\boldsymbol{\beta}})$ be? We then use an estimate of this variance upper bound to conduct valid hypothesis tests on $f(\boldsymbol{\beta})$.

To motivate this approach and provide a rationale for focusing on the variance upper-bound, consider testing:

$$H_0 : f(\boldsymbol{\beta}) \leq k \quad \text{against} \quad H_1 : f(\boldsymbol{\beta}) > k. \tag{3.1}$$

When the variance τ^2 can be consistently estimated, standard t-tests control size and are uniformly most powerful. The difficulty arises when the correlations across effects (ρ_{ij}) are unknown and τ^2 cannot be estimated. In this case, the problem can be framed as hypothesis testing with nuisance parameters, ρ_{ij} for $i \neq j$. Finding the UMP test in this setting corresponds to identifying the least favorable distribution of the nuisance parameters (Theorem 3.8.1 in Romano and Lehmann, 2005).[2] While least favorable distributions are often challenging to characterize (see, e.g., Elliott, Müller and Watson, 2015), our setting is simplified by the fact that the nuisance parameters enter the distribution of $f(\hat{\boldsymbol{\beta}})$ only through its variance. Since power is minimized when variance is maximized, finding the least favorable distribution—and hence the uniformly most powerful test—corresponds to finding the correlation structure that maximizes τ^2. This provides the statistical rationale for focusing on the variance upper bound.

In Section 3.1, we consider the general setting where no additional structure is imposed on the correlation structure, so the variance upper bound is determined solely by the mathematical constraints of the correlation matrix. In Section 3.2, we specialize to cases where treatment assignment is either completely randomized or randomized conditional on observed covariates.

[2] A least favorable distribution is the distribution on the nuisance parameters under which the test performs the worst, or in other words, the distribution under which the probability of correctly rejecting a false null is smallest. If a test controls size and has good power even under this "worst-case" scenario, then it will perform at least as well under all other admissible distributions.

3.1 Worst-Case Inference

We can re-express Equation (2.1) in terms of the correlations ρ_{ij} as follows:

$$\tau^2 = \sum_{i=1}^{d} \left(\sigma_i \frac{\partial f(\boldsymbol{\beta})}{\partial \beta_i} \right)^2 + \sum_{\substack{i=1 \\ \{i,j: i \neq j\}}}^{d} \sum_{j=1}^{d} \rho_{ij}\, \sigma_i \sigma_j \frac{\partial f(\boldsymbol{\beta})}{\partial \beta_i} \frac{\partial f(\boldsymbol{\beta})}{\partial \beta_j}. \tag{3.2}$$

Since β_i and σ_i can be consistently estimated from the observed data, obtaining an upper bound for τ^2 amounts to maximizing Equation (3.2) with respect to ρ_{ij} for $i \neq j$, subject to certain constraints. We formulate it as the following convex optimization problem:

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

Constraint (C.1) requires the variance-covariance matrix to be positive semidefinite; Constraint (C.2) enforces symmetry of the variance-covariance matrix; and Constraint (C.3) ensures that all pairwise correlations lie within $[-1, 1]$. Problem (**SDP.1**) is therefore a well-defined semidefinite program (SDP) that can be solved using existing optimization tools (Grant and Boyd, 2008, 2014). A key advantage of this formulation is its flexibility: we can incorporate available information about the correlations as additional constraints. For instance, in some cases, it may be known that two estimates are uncorrelated, such as when they are constructed using independent, non-overlapping samples. This information can be incorporated into the optimization problem by fixing the corresponding correlation to be zero. This flexibility is particularly important for the analysis in Section 3.2, where we leverage our characterization of the off-diagonal entries of the covariance matrix to impose theory-motivated sign restrictions.

We denote the maximum variance obtained by solving **SDP.1** as $\tau^2_{\max}$ and note the following. First, confidence intervals constructed using $\tau_{\max}$ will, by construction, have weakly higher coverage probability than those based on τ. While this reduces power, it guarantees size control; coverage is exact only when $\tau_{\max} = \tau$. Second, in settings where estimating the covariance across estimates is feasible but costly, we recommend that researchers first test their hypotheses using $\tau_{\max}$. Rejecting a null hypothesis under the worst-case variance implies that the null would also be rejected using the true variance. This allows researchers to conduct valid inference while avoiding the costs of computing the full covariance matrix. Finally, in Appendix Section A.1, we compare our approach to that of Cocci and Plagborg-Møller (2024) who propose a method for worst-case inference when matching structural parameters to empirical moments in overidentified settings. We show how tighter bounds can be found than what is implied by Lemma 1 in Cocci and Plagborg-Møller (2024), and illustrate through an example why maximizing the variance is challenging even in the simple case where there are no additional constraints beyond those in **SDP.1**.

3.2 Worst-Case Inference Under Random Treatment Assignment

In contrast to the generic case considered in Section 3.1, this section leverages information about treatment assignment to derive more powerful tests on cost-effectiveness parameters. When treatment is randomized—either completely or conditional on observables—we obtain a novel, interpretable characterization of the covariance matrix. This characterization allows researchers to impose sign or independence restrictions on the correlation matrix, grounded in theory, prior evidence, or auxiliary data. Incorporating these restrictions can substantially sharpen variance bounds and deliver more precise inference on policy-relevant parameters.

Let $Y_{ij}(1)$ denote the treated potential outcome j for unit i and $Y_{ij}(0)$ denote the control potential outcome j for unit i, where $j \in \{1, ..., d\}$. Let $Z_i \in \{0, 1\}$ indicate treatment assignment, where $Z_i = 1$ if unit i is treated and $Z_i = 0$ otherwise. We assume random assignment, meaning that treatment is independent of the full vector of potential outcomes:

$$\Big(Y_{ij}(1), Y_{ij}(0)\Big) \perp Z_i \quad \text{for all} \quad j = 1, \ldots, d$$

The observed outcome is $Y_{ij} = Z_i \cdot Y_{ij}(1) + (1 - Z_i) \cdot Y_{ij}(0)$. For each outcome j, the average treatment effect (ATE) is given by $\beta_j = \mathbb{E}[Y_{ij}(1) - Y_{ij}(0)]$, and we estimate it using the difference in sample means between the treated and control groups:

$$\widehat{\beta}_j = \frac{1}{n_1} \sum_{i:Z_i=1} Y_{i,j} - \frac{1}{n_0} \sum_{i:Z_i=0} Y_{i,j}$$

where n_1 and n_0 are the number of treated and control units, respectively. Let $\hat{\boldsymbol{\beta}} \in \mathbb{R}^d$ denote the vector of estimated treatment effects. In this setting, the asymptotic variance-covariance matrix $\mathbf{V}$ has a structure that allows for a simple and interpretable characterization, summarized in the following proposition.

Proposition 1. *Let β_p and β_q denote the average treatment effects of a randomized treatment $Z_i \in \{0, 1\}$ on two outcomes Y_{ip} and Y_{iq} respectively. Let the vector $\{(Y_{ip}(0), Y_{ip}(1), Y_{iq}(0), Y_{iq}(1), Z_i)\}_{i=1}^n$ be i.i.d. across units. Then, the asymptotic covariance, denoted by* AsyCov$(\cdot)$*, between the difference-in-means estimators $\widehat{\beta}_p$ and $\widehat{\beta}_q$ is given by:*

$$\text{AsyCov}\left(\widehat{\beta}_p, \widehat{\beta}_q\right) = \frac{\mathbb{C}\text{ov}(Y_{ip}, Y_{iq} \mid Z_i = 1)}{\mathbb{P}(Z_i = 1)} + \frac{\mathbb{C}\text{ov}(Y_{ip}, Y_{iq} \mid Z_i = 0)}{\mathbb{P}(Z_i = 0)}.$$

The proof of Proposition 1 is in Appendix Section B.1. The proposition shows that the asymptotic covariance between estimated treatment effects $\widehat{\beta}_p$ and $\widehat{\beta}_q$ depends only on the covariances of outcomes Y_{ip} and Y_{iq} within the treatment and control groups. In particular, if the outcomes are positively correlated within both groups, the treatment effects on those outcomes must also move in the same direction. For example, in the case of a randomized tax credit expansion called Paycheck Plus, the MVPF depends on the effects of the program on after-tax income, earnings, and labor force participation.[3] Since individuals with higher earnings also tend to have weakly higher after-tax income and are weakly more likely to participate in the labor force, Proposition 1 implies that the off-diagonal entries of $\mathbf{V}$ are non-negative. This information can be incorporated as additional constraints in the optimization problem **SDP.1**, thereby producing a (weakly) tighter upper bound. If it is known that all covariances between outcome pairs are non-negative as in the Paycheck Plus MVPF, we can solve the following optimization problem to obtain the variance upper bound:[4]

$$\begin{aligned}
&\underset{\{\rho_{ij}\}_{i,j=1}^{d}}{\text{Maximize}} && \tau^2 && && \textbf{SDP.2}\\
&\text{subject to} && \mathbf{V} \succeq 0 && && \text{(C.1)}\\
& && \rho_{ij} = \rho_{ji} && \forall\ i, j = 1, ..., d && \text{(C.2)}\\
& && \rho_{ij} \in [-1, 1] && \forall\ i, j = 1, ..., d && \text{(C.3)}\\
& && \rho_{ii} = 1 && \forall\ i = 1, ..., d && \text{(C.4)}\\
& && \rho_{ij} \geq 0 && \forall\ i, j = 1, ..., d && \text{(C.5)}
\end{aligned}$$

In Section 5, we show that adding Constraint C.5 to the optimization problem reduces the width of the Paycheck Plus MVPF confidence intervals by nearly 30%.

We also extend Proposition 1 to settings where treatment assignment is random only conditional on covariates, such that

$$(Y_{ij}(1), Y_{ij}(0)) \perp Z_i \mid \mathbf{X}_i \quad \text{for all} \quad j = 1, \ldots, d$$

[3] The Paycheck Plus program is studied in Miller et al. (2017) and the MVPF for this program is computed in Hendren and Sprung-Keyser (2020).

[4] The variance upper bounds implied by solving **SDP.1** and **SDP.2** are sharp, in the specific sense that these bounds are attainable: there exists a valid correlation matrix satisfying all imposed constraints (i.e., positive semi-definiteness and any sign restrictions) that generates exactly this variance. If additional information further constraining the set of feasible correlation matrices were available but not incorporated into the optimization problem, the resulting bound would remain valid but would be conservative.

where $\mathbf{X}_i \in \mathbb{R}^k$ is a vector of observed covariates. A similar characterization of the asymptotic covariance under unconfoundedness is provided in Appendix Section B.2.

4 Breakdown Analysis

In Section 3, we proposed a method that constructs worst-case confidence intervals for cost-effectiveness parameters, guaranteeing appropriate coverage regardless of the true correlation structure across causal effects. The strength of this method is its robustness: it delivers valid inference without making parametric assumptions about which correlation structures are more likely than others. The drawback is that tests based on the worst-case variance may leave policymakers underpowered to reject relevant null hypotheses. In this section, we develop a complementary approach to inference: we specify a set of plausible correlation structures, place a probability distribution over them, and ask: how likely is it that a given hypothesis would be rejected if the true correlation structure were drawn from this set?

We begin by specifying three elements. First, we fix the null hypothesis of interest. For example, a policymaker may wish to test whether a policy with positive willingness-to-pay "pays for itself," i.e., $H_0 : \text{MVPF} < 0$ vs. $H_1 : \text{MVPF} \geq 0$ (Bergstrom, Dodds and Rios, 2025). Second, we specify the set of admissible correlation structures. For instance, Proposition 1 may imply that correlations across the estimated causal effects are non-negative, so the admissible set is all correlation matrices with non-negative off-diagonal elements. Finally, we specify a probability distribution over this admissible set. For example, a policymaker might assume that all correlation structures in the admissible set are equally plausible. Alternatively, they might want to assume that correlation structures closer to independence are more plausible in their setting. We operationalize this by placing an LKJ prior (Lewandowski, Kurowicka and Joe, 2009) over the admissible set. The LKJ distribution has density $\pi(\rho) \propto \det(\rho)^{\eta-1}$, where η is the parameter governing which correlation structures are more likely than others. When $\eta = 1$, the prior is uniform over all correlation matrices. Larger values of η place more mass near the identity matrix, favoring weaker correlations.

Next, we repeatedly draw from the specified distribution of correlation structures and test whether the null hypothesis is rejected under each draw. We define the *breakdown statistic* as the share of correlation structures under which we are unable to reject the null hypothesis. For policymakers, the breakdown statistic provides a transparent measure of how fragile a conclusion is to uncertainty about correlations. A breakdown statistic close to zero implies that the conclusion is robust to most correlation structures, whereas a value close to one indicates that the null hypothesis is unlikely to be rejected under any plausible correlation structure. We refer to this approach of assessing how easily a conclusion "breaks down" under alternative correlation structures as breakdown analysis.[5] Finally, note that if a null hypothesis can be rejected even under the worst-case correlation structure derived in Section 3.1, the breakdown statistic must equal zero: by definition, all other admissible correlation structures imply a (weakly) smaller asymptotic variance than the worst case and therefore also lead to rejection of the null.

The exact algorithm for estimating the breakdown statistic is described in Appendix Section C.1; we provide a sketch of the algorithm here. We aim to assess the robustness of inference on $f(\boldsymbol{\beta})$ to uncertainty about the asymptotic correlation structure of the estimated causal effects $\hat{\boldsymbol{\beta}} \in \mathbb{R}^d$. We define the robust region RR_f as the set of admissible correlation matrices under which the null hypothesis $H_0 : f(\boldsymbol{\beta}) < k$ is rejected at level α:

$$\mathrm{RR}_f = \left\{ \rho \in \mathcal{R} : f(\hat{\boldsymbol{\beta}}) - z_\alpha \cdot \tau(\rho) \geq k \right\},$$

where $\tau^2(\rho)$ denotes the asymptotic variance of $f(\hat{\boldsymbol{\beta}})$ under the correlation matrix ρ, $\mathcal{R}$ is the set of all admissible correlation matrices, and z_α is the $1 - \alpha$ quantile of the standard normal distribution. We then define the breakdown statistic as the probability that the null is not rejected under an LKJ prior π on ρ:

$$\mathrm{BR}_f = 1 - \Pr_{\rho \sim \pi} \left[\rho \in \mathrm{RR}_f \right].$$

[5] See Manski and Pepper (2018); Masten and Poirier (2020); Diegert, Masten and Poirier (2022); Rambachan and Roth (2023); Spini (2024) for similar approaches.

To estimate the Breakdown Statistic, we sample $\rho^{(1)}, \ldots, \rho^{(N)}$ from the specified LKJ prior distribution π. For each draw of a correlation matrix $\rho^{(m)}$, where $m = 1, \ldots, N$, we compute the implied standard error $\tau^{(m)}$ and determine whether the null hypothesis is rejected. The estimated Breakdown Statistic is the proportion of draws under which we are unable to reject the null hypothesis of interest:

$$\widehat{\mathrm{BR}}_f = 1 - \frac{1}{N} \sum_{m=1}^{N} R^{(m)}.$$

5 Application: Marginal Value of Public Funds

We illustrate our method by conducting inference on the Marginal Value of Public Funds (MVPF). Hendren and Sprung-Keyser (2020) popularized the MVPF as a unified metric for evaluating the "bang-for-the-buck" of public spending. An MVPF of 1 means that a policy delivers one dollar of benefits to recipients for each dollar of net government cost. Formally, the MVPF is defined as the benefits provided to recipients of a policy divided by the net cost borne by the government:

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

where ΔW denotes the estimated benefits to individuals, ΔE is the government's initial expenditure on the policy, and ΔC is the estimated reduction in government costs induced by the policy's causal effects.

Three features of the MVPF framework make our proposed method particularly well suited for valid inference. First, the MVPF is a non-linear function of multiple causal effects. To illustrate, consider the MVPF of the expanded Earned Income Tax Credit (EITC) program, Paycheck Plus. Miller et al. (2017) estimate the causal effects of the program on several outcomes, including earnings, employment, and after-tax income. These estimates, reported in Table 1, form the input vector:[6]

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

From these causal effects, the MVPF for Paycheck Plus is constructed as:[7]

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

Second, in most applications, the only available information are the reported causal effect estimates and their standard errors. For example, the effects of Paycheck Plus are estimated using confidential administrative tax data, and the original study does not report the correlation structure across outcomes. Thus, the information available for inference is limited to the estimates and standard errors in Table 1. To conduct inference in this setting, Hendren and Sprung-Keyser (2020) assume a correlation structure across estimates. As we illustrate in Appendix Section A.2, relying on an assumed correlation structure can imply confidence intervals that are not guaranteed to have the correct coverage.

Third, Hendren and Sprung-Keyser (2020) show that reallocating spending from Policy B to Policy A is welfare-improving if and only if $MVPF_A > MVPF_B$. Testing the hypothesis $H_0 : MVPF_A \leq MVPF_B$, then, is central to the policy choice problem.[8] Our method provides a test for this hypothesis that controls size under any correlation structure.

We apply our inference method to the MVPF of eight government policies chosen to span different domains of public expenditure: three job-training programs (Job Start, Work

[6] $\widehat{\beta}_1$ and $\widehat{\beta}_2$ represent the estimated causal effect of Paycheck Plus on extensive margin labor supply in 2014 and 2015, respectively. $\widehat{\beta}_3$ and $\widehat{\beta}_5$ represent the estimated causal effect of Paycheck Plus on after-tax income in 2014 and 2015, respectively. $\widehat{\beta}_4$ and $\widehat{\beta}_6$ represent the estimated causal effect of Paycheck Plus on earnings in 2014 and 2015, respectively.

[7] Willingness-to-pay for Paycheck Plus is measured by the average bonus amount multiplied by the share of recipients who would have received the transfer absent labor market responses, while net cost is computed as the increase in after-tax income minus the increase in earnings, reflecting changes in tax payments and transfer spending.

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

Advance, Year Up), two cash transfers (Paycheck Plus, Alaska Universal Basic Income), a health insurance expansion (Medicare Part D), childcare spending (foster care provision), and an unemployment insurance (UI) expansion.[9] The estimated MVPFs and 95% confidence intervals constructed by solving **SDP.1** are shown in Figure 1. Details of each policy and its MVPF calculation are provided in Appendix Section D.

Several lessons emerge from Figure 1. First, even without assumptions on the off-diagonal entries of the variance-covariance matrix, we can reject the null that the MVPF of Job Start or Year Up exceeds one under any correlation structure, implying that a dollar spent on these job-training programs delivers less than a dollar in benefits. Second, using the variance upper bound, we test $H_0 : MVPF_{\text{Alaska UBI}} \leq MVPF_{\text{Job Start}}$.[10] We reject this null hypothesis, suggesting that reallocating spending from job-training programs to universal basic income programs could be welfare-enhancing, if the beneficiaries of both policies receive equal welfare weights. Finally, our estimates highlight meaningful statistical uncertainty in the relative ranking of some policies. For example, while the point estimates suggest that reallocating funds from job training to UI extensions is welfare-improving, our inference exercise shows that the uncertainty in these estimates precludes such a conclusion.

The only policy for which we have access to the underlying microdata is Medicare Part D. Using this data, we can recover the full variance-covariance matrix and compute exact confidence intervals, something that is infeasible for the other policies we study. Table 2 compares three sets of confidence intervals for the estimated MVPF of Medicare Part D: exact intervals using the estimated correlation structure, intervals assuming all causal effects are uncorrelated, and worst-case intervals from **SDP.1**. The exact confidence intervals rule out MVPF values below 0.80 and above 1.95, whereas the worst-case confidence intervals rule out values below 0.17 and above 2.57. These results highlight a key takeaway for practitioners: reporting the estimated covariance matrix across causal effects, when feasible, can

[9] The MVPFs for Job Start, Work Advance, Year Up, Paycheck Plus, and Alaska Universal Basic Income are computed in Hendren and Sprung-Keyser (2020). The MVPF for Medicare Part D is computed in Wettstein (2020). The MVPF for foster care provision is computed in Baron and Gross (2025). The MVPF for the UI expansion is computed in Huang and Yang (2021).

[10] Since these MVPFs are based on independent samples, we assume they are uncorrelated.

substantially improve the precision of ex-post inference.

A policymaker choosing among policies may care about whether we can robustly conclude that a policy "pays for itself," rather than focusing only on the statistical uncertainty in the estimated returns to each policy. To answer this question for policies with positive willingness-to-pay, we use the Breakdown approach described in Section 4.[11] The Breakdown Statistic measures robustness as the share of admissible correlation structures under which the null hypothesis H_0 : MVPF < 0 cannot be rejected. We compute this statistic using a uniform prior over the space of admissible correlation matrices.[12] Table 3 shows that H_0 : MVPF < 0 is not rejected under 13% of admissible correlation structures for Paycheck Plus, whereas it is rejected under all admissible correlation structures for the remaining policies.[13] This illustrates the value of the Breakdown Statistic: it shows not only whether a policy appears to pay for itself, but also how sensitive that conclusion is to assumptions about correlations across causal effects.

Finally, we turn to policies evaluated using randomized trials, the setting of interest in Section 3.2. In these cases, Proposition 1 provides an interpretable characterization of the covariance structure that allows us to impose sign restrictions on correlations across outcomes. For example, in the case of Paycheck Plus, it is plausible to assume that individuals with higher after-tax income also have higher earnings and are more likely to participate in the labor force. Incorporating such restrictions, we compute the MVPF confidence intervals by solving Problem **SDP.2**. Figure 2 reports the resulting intervals for Job Start, Paycheck Plus, Work Advance, and Year Up. A key takeaway is that sign restrictions can meaningfully sharpen inference: for Paycheck Plus, the data allow us to rule out MVPF values below -0.38 and above 2.37, reducing the width of the confidence interval by nearly 30% relative to the worst-case bound. However, these confidence intervals may still be conservative relative to the

[11]Here, we restrict attention to policies for which the lower bound of the worst-case 90% confidence interval for willingness-to-pay—the numerator of the MVPF—exceeds zero.

[12]In Appendix Section C.2, we replicate Table 3 for $\eta = 0.5, 2$.

[13]If the lower bound of the worst-case 90% confidence interval is larger than the policy-relevant threshold of $MVPF = 0$, we would reject the hypothesis H_0 : MVPF < 0 under all admissible correlation matrices. In that case, the Breakdown Statistic is mechanically equal to 0.

true underlying uncertainty in the MVPF estimates. In settings where auxiliary data sources (e.g., household survey data) are available, such information may be used to further restrict the admissible set of correlation matrices.[14] Taken together, this illustrates how progressively stronger assumptions about correlations can be used to transparently trade off robustness against worst-case correlation structures for increased precision in policy evaluation.

[14]We illustrate how auxiliary data can be leveraged to guide more restrictive assumptions on the correlation structure in Appendix Section D.1. Specifically, we compare confidence intervals for Paycheck Plus under four alternative assumptions about the correlation structure: (a) the worst case implied by **SDP.1**; (b) non-negative sign restrictions as implied by **SDP.2**; (c) additionally imposing that the correlation between after-tax income and earnings is at least as large as what is observed in the IPUMS-CPS (Flood et al., 2025); and (d) the assumption of perfect positive correlation across all causal effects considered by Hendren and Sprung-Keyser (2020).

References

Baron, E Jason, and Max Gross. 2025. "Is there a foster care-to-prison pipeline? Evidence from quasi-randomly assigned investigators." *Review of Economics and Statistics*, 1–46. 9, D.7, D.7

Bergstrom, Katy, William Dodds, and Juan Rios. 2025. "Optimal policy reforms." Pontificia Universidade Catolica do Rio de Janeiro (PUC-Rio), Departamento de Economia, Rio de Janeiro. 1, 4

Cave, George, et al. 1993. "JOBSTART. Final Report on a Program for School Dropouts." *New York: MDRC.* D.5

Chetty, Raj. 2009. "Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods." *Annu. Rev. Econ.*, 1(1): 451–488. 1

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

Cocci, Matthew D, and Mikkel Plagborg-Møller. 2024. "Standard errors for calibrated parameters." *Review of Economic Studies.* 1, 3.1, (document), A.1

Diegert, Paul, Matthew A Masten, and Alexandre Poirier. 2022. "Assessing omitted variable bias when the controls are endogenous." *arXiv preprint arXiv:2206.02303.* 5

Elliott, Graham, Ulrich K Müller, and Mark W Watson. 2015. "Nearly optimal tests when a nuisance parameter is present under the null hypothesis." *Econometrica*, 83(2): 771–811. 3

Fein, David, and Jill Hamadyk. 2018. "Bridging the opportunity divide for low-income youth: Implementation and early impacts of the year up program." *OPRE Report*, 65: 2018. D.4

Flood, Sarah, Miriam King, Renae Rodgers, Steven Ruggles, J. Robert Warren, Daniel Backman, Etienne Breton, Grace Cooper, Julia A. Rivera Drew, Stephanie Richards, and David Van Riper. 2025. "Integrated Public Use Microdata Series, Current Population Survey: Version 13.0 [dataset]." 14, D.1

Grant, Michael, and Stephen Boyd. 2014. "CVX: Matlab software for disciplined convex programming, version 2.1." 3.1

Grant, Michael C, and Stephen P Boyd. 2008. "Graph implementations for nonsmooth convex programs." 95–110, Springer. 3.1

Heckman, James J, Seong Hyeok Moon, Rodrigo Pinto, Peter A Savelyev, and Adam Yavitz. 2010. "The rate of return to the HighScope Perry Preschool Program." *Journal of public Economics*, 94(1-2): 114–128. 1

Hendra, Richard, David H Greenberg, Gayle Hamilton, Ari Oppenheim, Alexandra Pennington, Kelsey Schaberg, and Betsy L Tessler. 2016. "Encouraging evidence on a sector-focused advancement strategy: Two-year impacts from the WorkAdvance demonstration." *New York: MDRC.* D.3

Hendren, Nathaniel, and Ben Sprung-Keyser. 2020. "A unified welfare analysis of government policies." *The Quarterly Journal of Economics*, 135(3): 1209–1318. 1, 3, 5, 9, 14, 1, (document), A.2, 15, D.1, D.2, D.3, D.4, D.5

Huang, Po-Chun, and Tzu-Ting Yang. 2021. "The welfare effects of extending unemployment benefits: Evidence from re-employment and unemployment transfers." *Journal of Public Economics*, 202: 104500. 9, D.8, D.8

Jones, Damon, and Ioana Marinescu. 2022. "The labor market impacts of universal and permanent cash transfers: Evidence from the Alaska Permanent Fund." *American Economic Journal: Economic Policy*, 14(2): 315–340. D.2

Lewandowski, Daniel, Dorota Kurowicka, and Harry Joe. 2009. "Generating random correlation matrices based on vines and extended onion method." *Journal of multivariate analysis*, 100(9): 1989–2001. 4, 3

Manski, Charles F, and John V Pepper. 2018. "How do right-to-carry laws affect crime rates? Coping with ambiguity using bounded-variation assumptions." *Review of Economics and Statistics*, 100(2): 232–244. 5

Masten, Matthew A, and Alexandre Poirier. 2020. "Inference on breakdown frontiers." *Quantitative Economics*, 11(1): 41–111. 5

Miller, Cynthia, Lawrence F Katz, Gilda Azurdia, Adam Isen, and Caroline 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.* 3, 5, 1, D.1

Rambachan, Ashesh, and Jonathan Roth. 2023. "A more credible approach to parallel trends." *Review of Economic Studies*, 90(5): 2555–2591. 5

Romano, Joseph P, and EL Lehmann. 2005. "Testing statistical hypotheses." 3

Rose, Evan K. 2018. "The effects of job loss on crime: evidence from administrative data." *Available at SSRN 2991317.* 1

Ruggles, Steven, Catherine A Fitch, and Evan Roberts. 2018. "Historical census record linkage." *Annual review of sociology*, 44(1): 19–37. 1

Schaberg, Kelsey. 2017. "Can Sector Strategies Promote Longer-Term Effects? Three-Year Impacts from the WorkAdvance Demonstration." *MDRC.* D.3

Spini, Pietro Emilio. 2024. "Robustness, heterogeneous treatment effects and covariate shifts." *arXiv preprint arXiv:2112.09259.* 5

van der Vaart, Aad W. 1998. *Asymptotic Statistics.* Cambridge University Press. 1

Wettstein, Gal. 2020. "Retirement lock and prescription drug insurance: Evidence from medicare part d." *American Economic Journal: Economic Policy*, 12(1): 389–417. 9, 2, D.6, D.6

Zellner, Arnold. 1962. "An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias." *Journal of the American statistical Association*, 57(298): 348–368. 1, 2

Table 1: MVPF Calculation for Paycheck Plus

	(1)	(2)	(3)
	Year	Estimate	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 ($\widehat{\beta}_1$)	2014	0.90%	0.94%
Extensive Margin Labor Market ($\widehat{\beta}_2$)	2015	2.5%	1.00%
Impact on After Tax Income ($\widehat{\beta}_3$)	2014	654	198.75
Impact on Earnings ($\widehat{\beta}_4$)	2014	33	245.35
Impact on After Tax Income ($\widehat{\beta}_5$)	2015	645	265.17
Impact on Earnings ($\widehat{\beta}_6$)	2015	192	329.44
WTP		1071	
Net Government Costs		1074	
MVPF		0.996	

Notes: The table reports the inputs to compute the MVPF for the Paycheck Plus program. The causal effects and their corresponding standard errors are reported in Miller et al. (2017). Using these estimates as inputs, the MVPF for Paycheck Plus is computed in Hendren and Sprung-Keyser (2020).

Table 2: Inference for Medicare Part D MVPF

(1)	(2)	(3)	(4)
MVPF	Exact CI	Independence CI	Worst-Case CI
1.37	[0.80, 1.95]	[0.52, 2.22]	[0.18, 2.57]

Notes: The table reports 95% confidence intervals for the MVPF of the introduction of Medicare Part D, using causal effects reported in Wettstein (2020). When constructing the MVPF, we use the unconditional version of the estimated causal effects of the policy on income and labor force participation for simplicity. The exact approach through which it is computed is detailed in Appendix Section D. Column 1 reports the point estimate for the MVPF. Column 2 reports the exact confidence intervals for the estimated MVPF. The exact confidence intervals are computed with the Seemingly Unrelated Regression (SUR) approach of Zellner (1962) using the (publicly available) microdata underlying the causal effects in Wettstein (2020). Column 3 reports the confidence intervals under the assumption that all the causal effects are uncorrelated with each other, i.e., the off-diagonal entries of the variance-covariance matrix are equal to 0. Column 4 reports the confidence intervals computed by solving **SDP.1**, using the method described in Section 3.1.

Table 3: Breakdown Statistics for MVPF

	(1)	(2)
	Breakdown Statistic	Worst-Case 90% CI
Alaska UBI	0	[0.79, 1.05]
Foster Care	0	[0.33, 4.12]
Medicare Part D	0	[0.37, 2.38]
UI Extension	0	[0.01, 4.03]
Work Advance	0	[0.32, 1.23]
Year Up	0	[0.36, 0.50]
Paycheck Plus	0.13	[-0.63, 2.62]

Notes: The table reports the Breakdown Statistic (Column 1) and the worst-case 90% confidence intervals (Column 2) for the policy's MVPF estimate. The construction of the MVPF for each policy is detailed in Appendix Section D. The confidence intervals are computed using the method described in Section 3.1, by solving **SDP.1**. The Breakdown Statistic is computed using the method described in Section 4 and is defined with respect to the null hypothesis H_0 : MVPF < 0. We draw correlation matrices from a uniform LKJ prior (LKJ distribution with $\eta = 1$) over the space of admissible correlation matrices and evaluate rejection using the critical value corresponding to a one-sided test with size 5%.

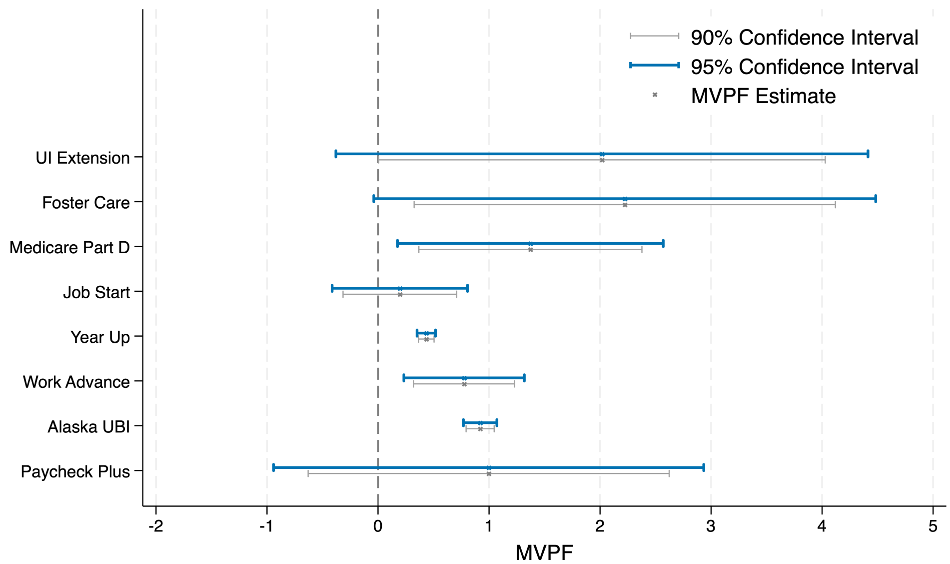

Figure 1: MVPF Worst-Case Confidence Intervals

Notes. The figure reports the worst-case 90% and 95% confidence intervals for the MVPF of eight different policies. The construction for the MVPF of each policy is detailed in Appendix Section D. The confidence intervals are computed using the method described in Section 3.1, by solving **SDP.1**.

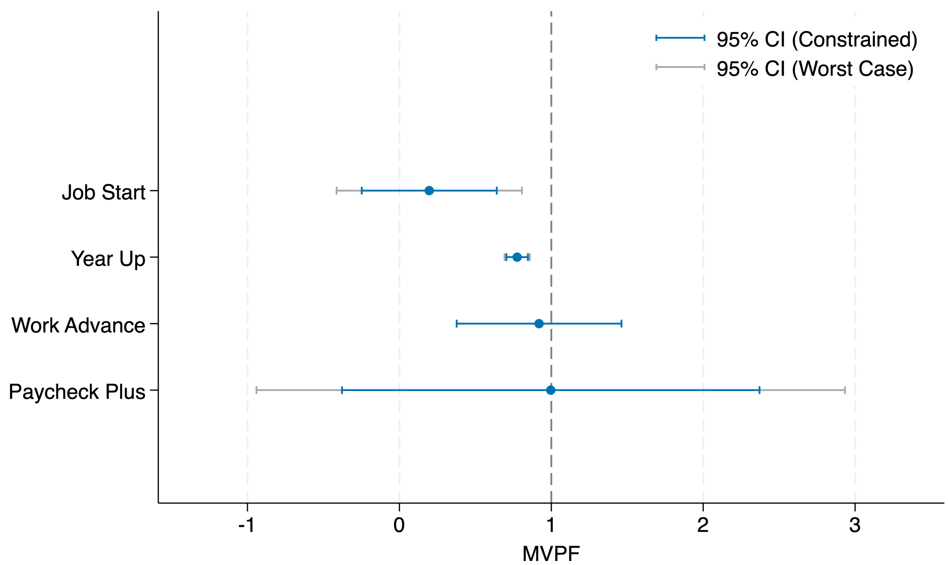

Figure 2: MVPF Confidence Intervals with Sign Constraints

Notes. The figure reports the 95% confidence intervals for the MVPF of four different policies that are evaluated using randomized trials. The confidence intervals are computed using the method described in Section 3.2, leveraging Proposition 1 to include sign constraints where appropriate. The construction for the MVPF of each policy as well as the sign constraints used are detailed in Appendix Section D.

Online Appendix

Table of Contents

A Connection to Existing Approaches

A.1 Inference Procedure in Cocci and Plagborg-Møller (2024)

In this section, we contrast our approach described in Section 3 with that of Cocci and Plagborg-Møller (2024), who study a related problem in the context of calibrating structural parameters to empirical moments in over-identified settings. They provide a convex optimization formulation for bounding the worst-case variance and, in Lemma 1, show that in the absence of additional restrictions, the variance can be maximized by inspecting the sign of the cross-partial term $\frac{\partial f(\boldsymbol{\beta})}{\partial \beta_i}\frac{\partial f(\boldsymbol{\beta})}{\partial \beta_j}$. If this product is positive, their result suggests setting $\rho_{ij} = 1$ maximizes the variance, while if it is negative, the variance is maximized by setting $\rho_{ij} = -1$.

To illustrate the challenging nature of worst-case inference, even in the absence of additional constraints, consider a stylized case where inference is conducted on a function of three causal effects ($d = 3$). Suppose additionally that

$$\frac{\partial f(\boldsymbol{\beta})}{\partial \beta_1}\frac{\partial f(\boldsymbol{\beta})}{\partial \beta_2} > 0, \quad \frac{\partial f(\boldsymbol{\beta})}{\partial \beta_1}\frac{\partial f(\boldsymbol{\beta})}{\partial \beta_3} > 0, \quad \frac{\partial f(\boldsymbol{\beta})}{\partial \beta_2}\frac{\partial f(\boldsymbol{\beta})}{\partial \beta_3} < 0.$$

Following Lemma 1, the implied "variance-maximizing" correlation matrix would be:

	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$
$\widehat{\beta}_1$	1	1	1
$\widehat{\beta}_2$	1	1	-1
$\widehat{\beta}_3$	1	-1	1

However, this matrix has an eigenvalue equal to -1, meaning it is not positive semidefinite and therefore does not satisfy Constraint (C.1) in **SDP.1**. While the bound implied by such a correlation structure is a valid upper bound, a tighter upper bound can be obtained by explicitly enforcing the positive semidefiniteness constraint, as in the convex optimization problem **SDP.1**. This example highlights the difficulty of identifying the tightest possible worst-case variance bound, even in the absence of additional constraints.

Finally, while Cocci and Plagborg-Møller (2024) note that additional constraints can be incorporated directly into their convex optimization framework, it may not always be obvious to the researcher what those constraints should be. Section 3.2 provides guidance on how to introduce such constraints in practice: for example, by exploiting structure in randomized treatment designs or by leveraging settings where treatment is random conditional on observables.

A.2 Inference Procedure in Hendren and Sprung-Keyser (2020)

To construct confidence intervals for Marginal Value of Public Funds (MVPF) estimates, Hendren and Sprung-Keyser (2020) adopt a parametric bootstrap procedure. They begin by specifying a correlation structure across the underlying causal effect estimates. This correlation structure is user-specified and chosen to "maximize the width of [their] confidence intervals where estimates are from the same sample."[15] Conditional on this specification, they repeatedly draw from a joint normal distribution centered at the reported estimates with the chosen correlation structure. For each draw, they compute the implied MVPF, generating a simulated distribution of the statistic. The 2.5th and 97.5th percentiles of this distribution are used to construct the confidence intervals.

If the correlation structure specified in the first step of their procedure happened to coincide with the one that maximizes the width of the confidence intervals, then the bootstrap approach of Hendren and Sprung-Keyser (2020) would yield valid inference, and the resulting intervals would match ours. The key distinction is that our method does not assume this structure ex ante: instead, we formally identify it by solving **SDP.1**. Because the variance-maximizing correlation structure is rarely obvious, simply positing one does not guarantee size control. By casting the search as an optimization problem, our approach ensures valid

[15]The method is described in detail in Online Appendix H of Hendren and Sprung-Keyser (2020) as well as Section I.A of the replication files, accessible at `https://github.com/OpportunityInsights/welfare_analysis`.

inference regardless of the true correlation structure.

To illustrate the pitfalls of assuming a candidate worst-case correlation structure, we revisit the MVPF for Paycheck Plus, described in Section 5. Hendren and Sprung-Keyser (2020) assume that the underlying causal effects are perfectly positively correlated, which yields confidence intervals for the MVPF of $[0.870, 1.190]$. However, solving Problem **SDP.1** reveals that the true worst-case correlation structure is instead

	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$	$\widehat{\beta}_4$	$\widehat{\beta}_5$	$\widehat{\beta}_6$
$\widehat{\beta}_1$	1					
$\widehat{\beta}_2$	1	1				
$\widehat{\beta}_3$	1	1	1			
$\widehat{\beta}_4$	−1	−1	−1	1		
$\widehat{\beta}_5$	1	1	1	−1	1	
$\widehat{\beta}_6$	−1	−1	−1	1	−1	1

where each entry denotes the pairwise correlation $\rho_{ij} = \mathrm{Corr}(\widehat{\beta}_i, \widehat{\beta}_j)$. Under the correct worst-case correlation shown above, the implied confidence intervals are $[-0.941, 2.934]$. This example highlights how assuming a correlation structure—even one designed to be conservative—need not deliver valid inference if it is misspecified. By contrast, our optimization-based approach guarantees size control by formally identifying the correlation structure that maximizes the variance.

A further advantage of our framework is that it avoids reliance on correlation matrices that may not be feasible. In practice, it can be difficult to tell whether a user-specified matrix respects the geometry of a valid correlation structure—specifically, whether it is positive semidefinite. As we show in Appendix Section A.1, imposing this constraint explicitly can lead to tighter confidence intervals with valid coverage rates.

B Proposition 1

B.1 Proof of Proposition 1

We begin by writing

$$\hat{\beta}_p = \bar{Y}_{p,1} - \bar{Y}_{p,0}, \quad \hat{\beta}_q = \bar{Y}_{q,1} - \bar{Y}_{q,0}$$

where

$$\bar{Y}_{p,1} = \frac{1}{n_1} \sum_{i:Z_i=1} Y_{ip}, \quad \bar{Y}_{p,0} = \frac{1}{n_0} \sum_{i:Z_i=0} Y_{ip}, \qquad \text{and similarly for } \bar{Y}_{q,1}, \bar{Y}_{q,0}$$

Then,

$$\begin{aligned}
\mathbb{C}\text{ov}\left(\hat{\beta}_p, \hat{\beta}_q\right) &= \mathbb{C}\text{ov}\left(\bar{Y}_{p,1} - \bar{Y}_{p,0}, \bar{Y}_{q,1} - \bar{Y}_{q,0}\right) \\
&= \mathbb{C}\text{ov}\left(\bar{Y}_{p,1}, \bar{Y}_{q,1}\right) + \mathbb{C}\text{ov}\left(\bar{Y}_{p,0}, \bar{Y}_{q,0}\right) - \mathbb{C}\text{ov}\left(\bar{Y}_{p,1}, \bar{Y}_{q,0}\right) - \mathbb{C}\text{ov}\left(\bar{Y}_{p,0}, \bar{Y}_{q,1}\right)
\end{aligned}$$

Under random assignment and i.i.d. sampling, the treated and control groups are independent samples from the population. Therefore,

$$\mathbb{C}\text{ov}\left(\bar{Y}_{p,1}, \bar{Y}_{q,0}\right) = 0, \quad \mathbb{C}\text{ov}\left(\bar{Y}_{p,0}, \bar{Y}_{q,1}\right) = 0$$

So:

$$\mathbb{C}\text{ov}\left(\hat{\beta}_p, \hat{\beta}_q\right) = \mathbb{C}\text{ov}\left(\bar{Y}_{p,1}, \bar{Y}_{q,1}\right) + \mathbb{C}\text{ov}\left(\bar{Y}_{p,0}, \bar{Y}_{q,0}\right)$$

We now characterize each term. Because $\{Y_{ip}, Y_{iq}\}_{i:Z_i=1}$ is an i.i.d. sample from the treated population of size n_1, we have:

$$\mathbb{C}\text{ov}\left(\bar{Y}_{p,1}, \bar{Y}_{q,1}\right) = \frac{1}{n_1} \mathbb{C}\text{ov}\left(Y_{ip}, Y_{iq} \mid Z_i = 1\right)$$

Similarly, for the control group:

$$\mathbb{C}\text{ov}\left(\bar{Y}_{p,0}, \bar{Y}_{q,0}\right) = \frac{1}{n_0} \mathbb{C}\text{ov}\left(Y_{ip}, Y_{iq} \mid Z_i = 0\right)$$

So,

$$\mathbb{C}\mathrm{ov}\left(\hat{\beta}_p, \hat{\beta}_q\right) = \frac{1}{n_1}\mathbb{C}\mathrm{ov}\left(Y_{ip}, Y_{iq} \mid Z_i = 1\right) + \frac{1}{n_0}\mathbb{C}\mathrm{ov}\left(Y_{ip}, Y_{iq} \mid Z_i = 0\right)$$

As $n \to \infty$, the Law of Large Numbers implies:

$$\frac{n_1}{n} \xrightarrow{p} \mathbb{P}(Z_i = 1), \quad \frac{n_0}{n} \xrightarrow{p} \mathbb{P}(Z_i = 0)$$

So,

$$\mathbb{C}\mathrm{ov}\left(\hat{\beta}_p, \hat{\beta}_q\right) = \frac{1}{n}\left(\frac{\mathbb{C}\mathrm{ov}\left(Y_{ip}, Y_{iq} \mid Z_i = 1\right)}{\mathbb{P}(Z_i = 1)} + \frac{\mathbb{C}\mathrm{ov}\left(Y_{ip}, Y_{iq} \mid Z_i = 0\right)}{\mathbb{P}(Z_i = 0)}\right) + o_p(1)$$

Multiplying both sides by n, we obtain the (p, q)-th entry of the asymptotic covariance matrix for $\sqrt{n}(\hat{\beta} - \beta)$ as:

$$\frac{\mathbb{C}\mathrm{ov}\left(Y_{ip}, Y_{iq} \mid Z_i = 1\right)}{\mathbb{P}(Z_i = 1)} + \frac{\mathbb{C}\mathrm{ov}\left(Y_{ip}, Y_{iq} \mid Z_i = 0\right)}{\mathbb{P}(Z_i = 0)}$$

which proves the proposition. $\square$

B.2 Extension of Proposition 1 to Unconfoundedness

In this section, we extend the result in Proposition 1 to the setting with covariates. Specifically, we relax the assumption in Proposition 1 that $\left(Y_{ij}(1), Y_{ij}(0)\right) \perp Z_i$ for all $j = 1, \ldots, d$, and instead assume that

$$(Y_{ij}(1), Y_{ij}(0)) \perp Z_i \mid \mathbf{X}_i \quad \text{for all} \quad j = 1, \ldots, d$$

where $\mathbf{X}_i \in \mathbb{R}^k$ is a vector of observed covariates.

Let the data $\{(Y_{ij}, Z_i, \mathbf{X}_i)\}_{i=1}^n$ be i.i.d. across units. In order to consistently estimate the average treatment effect under unconfoundedness using the oft-adopted regression estimator, we make two additional assumptions. First, we assume that there is overlap, such that the probability of being treated is bounded away from 0 and 1 at each covariate value in the support:

$$0 < \mathbb{P}\left(Z_i = 1 \mid \mathbf{X}_i\right) < 1 \quad \text{a.s.}$$

We also assume that the true conditional expectation function is linear in covariates. Specifically, we assume that, for each outcome j,

$$\mathbb{E}\left[Y_{ij} \mid Z_i, \mathbf{X}_i\right] = \alpha_j + \tau_j Z_i + \mathbf{X}_i^\top \gamma_j$$

Let $\hat{\tau}_j$ be the OLS coefficient on Z_i in a regression of Y_{ij} on Z_i and $\mathbf{X}_i$. In this section, we characterize the asymptotic covariance between the estimated treatment effects $\hat{\tau}_p$ and $\hat{\tau}_q$.

For each outcome $\boldsymbol{j}$, consider the linear regression:

$$Y_{ij} = \alpha_j + \tau_j Z_i + \mathbf{X}_i^\top \gamma_j + \varepsilon_{ij}$$

Define $\tilde{Z}_i := Z_i - \Pi_Z \mathbf{X}_i$, the residual from regressing Z_i on $\mathbf{X}_i$ and $\tilde{Y}_{ij} := Y_{ij} - \Pi_j \mathbf{X}_i$, the residual from regressing Y_{ij} on $\mathbf{X}_i$, where Π_Z and Π_j are the population projections. Then by the Frisch-Waugh-Lovell theorem, the coefficient $\hat{\tau}_j$ is equal to the slope coefficient in the regression of $\tilde{Y}_{ij}$ on $\tilde{Z}_i$, i.e.,

$$\hat{\tau}_j = \frac{\sum_{i=1}^n \tilde{Z}_i \tilde{Y}_{ij}}{\sum_{i=1}^n \tilde{Z}_i^2}$$

Define,

$$w_i := \frac{\tilde{Z}_i}{\sum_{j=1}^n \tilde{Z}_j^2}$$

Then,

$$\hat{\tau}_j = \sum_{i=1}^n w_i \tilde{Y}_{ij}$$

The asymptotic covariance between the average treatment effect estimators $\hat{\tau}_p$ and $\hat{\tau}_q$ is given by,

$$\operatorname{Cov}\left(\hat{\tau}_p, \hat{\tau}_q\right) = \operatorname{Cov}\left(\sum_{i=1}^n w_i \tilde{Y}_{ip}, \sum_{j=1}^n w_j \tilde{Y}_{jq}\right) = \sum_{i=1}^n w_i^2 \cdot \operatorname{Cov}\left(\tilde{Y}_{ip}, \tilde{Y}_{iq}\right) + \sum_{i \neq j} w_i w_j \cdot \operatorname{Cov}\left(\tilde{Y}_{ip}, \tilde{Y}_{jq}\right)$$

Under i.i.d. sampling, $\text{Cov}\left(\tilde{Y}_{ip}, \tilde{Y}_{jq}\right) = 0$ for $i \neq j$, so:

$$\text{Cov}\left(\hat{\tau}_p, \hat{\tau}_q\right) = \sum_{i=1}^{n} w_i^2 \cdot \text{Cov}\left(\tilde{Y}_{ip}, \tilde{Y}_{iq}\right)$$

By the Law of Large Numbers, $\frac{1}{n}\sum_{i=1}^{n} \tilde{Z}_i^2 \xrightarrow{p} \mathbb{E}\left[\tilde{Z}_i^2\right]$ and $\frac{1}{n}\sum_{i=1}^{n} \tilde{Z}_i^2 \cdot \text{Cov}\left(\tilde{Y}_{ip}, \tilde{Y}_{iq}\right) \xrightarrow{p} \mathbb{E}\left[\tilde{Z}_i^2 \cdot \text{Cov}\left(\tilde{Y}_{ip}, \tilde{Y}_{iq}\right)\right]$. Therefore, we obtain the (p, q)-th entry of the asymptotic covariance matrix for $\sqrt{n}(\hat{\tau} - \tau)$ as

$$\text{Cov}\left(\hat{\tau}_p, \hat{\tau}_q\right) = \frac{\mathbb{E}\left[\tilde{Z}_i^2 \cdot \text{Cov}\left(\tilde{Y}_{ip}, \tilde{Y}_{iq}\right)\right]}{\left(\mathbb{E}\left[\tilde{Z}_i^2\right]\right)^2}$$

Thus, if

$$\text{Cov}\left(Y_{ip}, Y_{iq} \mid \mathbf{X}_i, Z_i = 1\right) \geq 0 \quad \text{and} \quad \text{Cov}\left(Y_{ip}, Y_{iq} \mid \mathbf{X}_i, Z_i = 0\right) \geq 0 \quad \text{a.s.,}$$

then it follows that,

$$\text{Cov}\left(\hat{\tau}_p, \hat{\tau}_q\right) \geq 0$$

$\square$

C Breakdown Statistic Details

In this section, we describe a step-by-step procedure to operationalize the Breakdown Approach introduced in Section 4. To illustrate the Breakdown approach, we estimate the proportion of correlation structures under which we cannot reject the null hypothesis H_0 : MVPF < 0 for a subset of policies with positive willingness-to-pay in Section 5. Though we restrict our focus on this hypothesis, it is worth noting that the breakdown approach may be applicable more generally. For instance, one may be interested in knowing the share of correlation matrices for which we can reject the null that the policy does not have a positive willingness-to-pay.

C.1 Computation Algorithm

1. Fix a null hypothesis of interest:

$$H_0 : f(\boldsymbol{\beta}) < k \quad \text{against} \quad H_1 : f(\boldsymbol{\beta}) \geq k.$$

2. **Compute the estimate and its gradient.** Calculate $f(\hat{\boldsymbol{\beta}})$ and the gradient $\nabla f(\hat{\boldsymbol{\beta}})$.

3. **Draw correlation matrices.** Sample $\rho^{(1)}, \ldots, \rho^{(N)} \sim \pi$, where π is a prior distribution over the space of valid correlation matrices $\mathcal{R}$. We adopt the LKJ prior (Lewandowski, Kurowicka and Joe, 2009), which has density:

$$\pi(\rho) \propto \det(\rho)^{\eta-1}.$$

   When $\eta = 1$, the prior is uniform over $\mathcal{R}$. Larger values of η place more mass near the identity matrix, favoring weaker correlations.

4. **Test under each draw.** For each draw of a correlation matrix $\rho^{(m)}$, where $m = 1, \ldots, N$, compute the implied standard error $\tau^{(m)}$ and determine whether the null

hypothesis is rejected:

$$R^{(m)} = \mathbb{1}\left\{f(\hat{\boldsymbol{\beta}}) - z_\alpha \cdot \tau^{(m)} \geq k\right\},$$

where z_α is the $1 - \alpha$ quantile of the standard normal distribution.

5. **Estimate the breakdown statistic.** Compute:

$$\widehat{\mathrm{BR}}_f = 1 - \frac{1}{N}\sum_{m=1}^{N} R^{(m)}.$$

This statistic measures the proportion of correlation structures under which we cannot reject the null hypothesis, assuming $\rho \sim \pi$.

C.2 Varying η

Recall that the LKJ distribution has density $\pi(\rho) \propto \det(\rho)^{\eta-1}$, where η is the parameter governing which correlation structures are more likely than others. When $\eta = 1$, the prior is uniform over all admissible correlation matrices. Larger values of η place more mass near the identity matrix, favoring weaker correlations. In this section, we replicate the Breakdown Statistics reported in Table 3 for three values of η: 0.5, 1, and 2. In the case of Paycheck Plus, we are more likely to reject the null of $H_0 : MVPF < 0$ when the correlation matrices are more likely to be near the identity matrix, leading to lower Breakdown Statistics as η increases.

Table C.1: Breakdown Statistics for MVPF

	(1)	(2)	(3)
	η=0.5	η=1	η=2
Alaska UBI	0	0	0
Foster Care	0	0	0
Medicare Part D	0	0	0
UI Extension	0	0	0
Work Advance	0	0	0
Year Up	0	0	0
Paycheck Plus	0.15	0.13	0.10

Notes: The table replicates the results in Table 3 for $\eta = 0.5, 1, 2$.

D Policy Details

In this section, we detail the construction of the MVPF for all policies discussed in Section 5. To illuminate which correlations drive the worst-case results shown in Figure 1, for each policy we explicitly report: (a) the correlation matrix that solves **SDP.1**, and (b) the weight each correlation coefficient carries in the asymptotic variance of $f(\hat{\boldsymbol{\beta}})$. Recall from Equation 3.2 that the asymptotic variance of $f(\hat{\boldsymbol{\beta}})$ is given by:

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

We define the term $\sigma_i \sigma_j \frac{\partial f(\boldsymbol{\beta})}{\partial \beta_i} \frac{\partial f(\boldsymbol{\beta})}{\partial \beta_j}$ as the weight the correlation coefficient ρ_{ij} receives in the asymptotic variance of $f(\hat{\boldsymbol{\beta}})$, and report the sample analogue of this term for all $i, j \in \{1, \ldots, d\}$.

We defer further discussion of the underlying economic assumptions of each MVPF to the original source paper providing the construction.

D.1 Paycheck Plus

The estimates used to construct the MVPF for Paycheck Plus are drawn from Miller et al. (2017). The estimates are summarized in the following Table: We replicate the construction

Table D.1: MVPF Calculation for Paycheck Plus

	(1)	(2)	(3)
	Year	Estimate	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 ($\widehat{\beta}_1$)	2014	0.90%	0.94%
Extensive Margin Labor Market ($\widehat{\beta}_2$)	2015	2.5%	1.00%
Impact on After Tax Income ($\widehat{\beta}_3$)	2014	654	198.75
Impact on Earnings ($\widehat{\beta}_4$)	2014	33	245.35
Impact on After Tax Income ($\widehat{\beta}_5$)	2015	645	265.17
Impact on Earnings ($\widehat{\beta}_6$)	2015	192	329.44

of the MVPF for Paycheck Plus from Hendren and Sprung-Keyser (2020), as follows:

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

The correlation matrix that maximizes **SDP.1** is:

	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$	$\widehat{\beta}_4$	$\widehat{\beta}_5$	$\widehat{\beta}_6$
$\widehat{\beta}_1$						
$\widehat{\beta}_2$	1					
$\widehat{\beta}_3$	1	1				
$\widehat{\beta}_4$	-1	-1	-1			
$\widehat{\beta}_5$	1	1	1	-1		
$\widehat{\beta}_6$	-1	-1	-1	1	-1	

Below, we report the estimated weights associated with the corresponding correlation coefficient ρ_{ij} in the asymptotic variance expression of $f(\hat{\boldsymbol{\beta}})$ derived in Equation 3.2. For instance, the value corresponding to $\widehat{\beta}_1$ (Column 1) and $\widehat{\beta}_2$ (Row 2) is equal to $\hat{\sigma}_1\hat{\sigma}_2\frac{\partial f(\hat{\boldsymbol{\beta}})}{\partial \beta_1}\frac{\partial f(\hat{\boldsymbol{\beta}})}{\partial \beta_2}$:

	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$	$\widehat{\beta}_4$	$\widehat{\beta}_5$	$\widehat{\beta}_6$
$\widehat{\beta}_1$						
$\widehat{\beta}_2$	0.00015					
$\widehat{\beta}_3$	0.00226	0.00233				
$\widehat{\beta}_4$	−0.00278	−0.00288	−0.04197			
$\widehat{\beta}_5$	0.00301	0.00311	0.04536	−0.05599		
$\widehat{\beta}_6$	−0.00374	−0.00386	−0.05635	0.06956	−0.07518	

In Figure D.1, we compare confidence intervals for the MVPF of Paycheck Plus under a sequence of increasingly restrictive assumptions on the asymptotic correlation structure across causal effects.

The top interval ("Worst-Case") reports the maximally conservative confidence interval obtained by solving **SDP.1**, which allows the correlation matrix to vary freely over all positive semidefinite matrices that are consistent with the reported standard errors. This interval guarantees valid inference uniformly over all admissible correlation structures, but may be overly conservative.

The second interval ("Sign Restrictions") corresponds to the assumptions imposed in **SDP.2**, where we restrict the correlation across all causal effects to be non-negative. This restriction is motivated by our result in Proposition 1, which implies that if outcomes are positively correlated in both the treatment and control groups, the resulting causal effects will also be positively correlated. While the assumption that all causal effects are (weakly) positively correlated meaningfully sharpens inference, the implied confidence intervals may still be conservative if, for instance, the correlation between after-tax income and earnings is close to +1.

The third interval ("CPS Benchmark") further restricts the admissible set of correlation matrices by restricting the correlation between the effect of the policy on after-tax income and earnings to be at least as large as 0.85, the correlation between after-tax income and earnings observed in the IPUMS-CPS for the corresponding year (Flood et al., 2025).[16] While this assumption further sharpens inference, the resulting confidence intervals may fail to achieve nominal coverage if the correlation between the effect of Paycheck Plus on after-tax income and earnings is smaller than 0.85.

The final interval ("HSK (2020)") reproduces the assumption in Hendren and Sprung-Keyser (2020) that all causal effects are perfectly positively correlated. Again, the resulting confidence interval need not achieve nominal coverage if the true correlation structure deviates from perfect positive correlation.

Taken together, Figure D.1 illustrates that moving from worst-case bounds to theory- and data-informed restrictions yields progressively more informative confidence intervals, at the cost of relying on stronger assumptions about the underlying correlation structure.

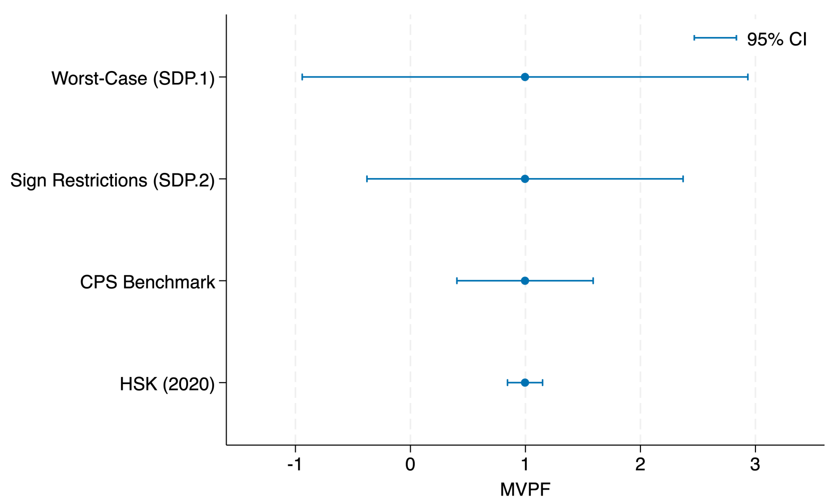

Figure D.1: Paycheck Plus MVPF CIs Under Alternative Correlation Assumptions

[16]The value 0.85 corresponds to the correlation between after-tax income and earnings estimated in the 2014 and 2015 Current Population Survey Annual Social and Economic Supplement (CPS ASEC), accessed via IPUMS. We restrict the sample to household heads with non-missing income and tax variables, construct earnings as the sum of wage, business, and farm income, and construct after-tax income as total income net of federal income taxes, state income taxes, and payroll taxes.

D.2 Alaska UBI

The estimates used to construct the MVPF for Alaska UBI are drawn from Jones and Marinescu (2022). We replicate the construction of the MVPF for Alaska UBI from Hendren and Sprung-Keyser (2020), as follows:

Table D.2: MVPF Calculation for Alaska UBI

	(1) Estimate	(2) SE
Full-Time Employment Effect ($\widehat{\beta}_1$)	0.001	0.016
Part-Time Employment Effect ($\widehat{\beta}_2$)	0.018	0.007

$$MVPF_{\text{Alaska UBI}} = f(\widehat{\beta}) = \frac{1000}{1000 - \left(\widehat{\beta}_1 \times 5567.88 \times \frac{1000}{1602}\right) + \left(0.2 \times 0.5 \times \widehat{\beta}_2 \times \frac{1000}{1602} \times 80830.57\right)}$$
$$= 0.92.$$

SDP.1 is maximized when the asymptotic correlation between $\widehat{\beta}_1$ and $\widehat{\beta}_2$ is equal to -1, and the estimated weight associated with the correlation coefficient ρ_{12} in the asymptotic variance expression of $f(\hat{\boldsymbol{\beta}})$, denoted by $\hat{\sigma}_1\hat{\sigma}_2\frac{\partial f(\hat{\boldsymbol{\beta}})}{\partial\beta_1}\frac{\partial f(\hat{\boldsymbol{\beta}})}{\partial\beta_2}$, is equal to -0.00142

D.3 Work Advance

The estimates used to construct the MVPF for Work Advance are drawn from Hendra et al. (2016) and Schaberg (2017). We replicate the construction of the MVPF for Work Advance

Table D.3: MVPF Calculation for Work Advance

	(1)	(2)
	Estimate	SE
Year 2 Earnings Effect ($\widehat{\beta}_1$)	1945	692.90
Year 3 Earnings Effect ($\widehat{\beta}_2$)	1865	664.40

from Hendren and Sprung-Keyser (2020), as follows:

$$MVPF_{\text{Work Advance}} = f(\widehat{\beta})\frac{\frac{\widehat{\beta}_1 \times (1-0.003)}{1.03} + \frac{\widehat{\beta}_2 \times (1-0.003)}{1.03^2}}{5641 - 940 - \widehat{\beta}_1 \times 0.003 - \widehat{\beta}_2 \times 0.003}$$
$$= 0.78$$

SDP.1 is maximized when the asymptotic correlation between $\widehat{\beta}_1$ and $\widehat{\beta}_2$ is equal to 1, and the estimated weight associated with the correlation coefficient ρ_{12} in the asymptotic variance expression of $f(\hat{\boldsymbol{\beta}})$, denoted by $\hat{\sigma}_1\hat{\sigma}_2\frac{\partial f(\hat{\boldsymbol{\beta}})}{\partial \beta_1}\frac{\partial f(\hat{\boldsymbol{\beta}})}{\partial \beta_2}$, is equal to 0.01914. In Figure 2, we assume that the asymptotic correlation across both causal effects is non-negative.

D.4 Year Up

The estimates used to construct the MVPF for Year Up are drawn from Fein and Hamadyk (2018). We replicate the construction of the MVPF for Year Up from Hendren and Sprung-

Table D.4: Year Up

	(1)	(2)
	Estimate	SE
Year 0 Earnings ($\widehat{\beta}_1$)	-5338	238
Year 1 Earnings ($\widehat{\beta}_2$)	5181	474
Year 2 Earnings ($\widehat{\beta}_3$)	7011	619
Discount Rate	3%	
Tax Rate	18.6%	
Per-Participant Cost	$28,290	
Student Stipend	$6,614	

Keyser (2020), as follows:

$$
\begin{aligned}
MVPF_{\text{Year Up}} = f(\widehat{\beta}) &= \frac{(1-0.186)\times(\widehat{\beta}_1+\widehat{\beta}_2/0.03+\widehat{\beta}_3/1.03^2)+6614}{28290-0.186\times(\widehat{\beta}_1+\widehat{\beta}_2+\widehat{\beta}_3)} \\
&= 0.43
\end{aligned}
$$

The correlation matrix that maximizes SDP.1 is:

	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$
$\widehat{\beta}_1$			
$\widehat{\beta}_2$	1		
$\widehat{\beta}_3$	1	1	

Below, we report the estimated weights associated with the corresponding correlation coefficient ρ_{ij} in the asymptotic variance expression of $f(\hat{\boldsymbol{\beta}})$ derived in Equation 3.2. For instance, the value corresponding to $\widehat{\beta}_1$ (Column 1) and $\widehat{\beta}_2$ (Row 2) is equal to $\hat{\sigma}_1\hat{\sigma}_2\frac{\partial f(\hat{\boldsymbol{\beta}})}{\partial\beta_1}\frac{\partial f(\hat{\boldsymbol{\beta}})}{\partial\beta_2}$:

	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$
$\widehat{\beta}_1$			
$\widehat{\beta}_2$	0.00012		
$\widehat{\beta}_3$	0.00015	0.00030	

In Figure 2, we assume that the asymptotic correlation across $\widehat{\beta}_2$ and $\widehat{\beta}_3$ is non-negative, and the correlation of $\widehat{\beta}_1$ with both $\widehat{\beta}_2$ and $\widehat{\beta}_3$ is non-positive.

D.5 Job Start

The estimates used to construct the MVPF for Job Start are drawn from Cave et al. (1993). We replicate the construction of the MVPF for Job Start from Hendren and Sprung-Keyser

Table D.5: MVPF Calculation for Job Start

	(1)	(2)
	Estimate	SE
Year 1 Earnings Effect ($\widehat{\beta}_1$)	-499	151.65
Year 2 Earnings Effect ($\widehat{\beta}_2$)	-121	209.20
Year 3 Earnings Effect ($\widehat{\beta}_3$)	423	258.67
Year 4 Earnings Effect ($\widehat{\beta}_4$)	410	267.25
Year 1 AFDC Effect ($\widehat{\beta}_5$)	63	53.96
Year 2 AFDC Effect ($\widehat{\beta}_6$)	24	62.94
Year 3 AFDC Effect ($\widehat{\beta}_7$)	-3	85.47
Year 4 AFDC Effect ($\widehat{\beta}_8$)	-11	84.97
Year 1 Food Stamps Effect ($\widehat{\beta}_9$)	-45	35.66
Year 2 Food Stamps Effect ($\widehat{\beta}_{10}$)	-42	34.83
Year 3 Food Stamps Effect ($\widehat{\beta}_{11}$)	31	40.94
Year 4 Food Stamps Effect ($\widehat{\beta}_{12}$)	31	45.21
Year 1 General Assistance Effect ($\widehat{\beta}_{13}$)	24	23.54
Year 2 General Assistance Effect ($\widehat{\beta}_{14}$)	7	15.14
Year 3 General Assistance Effect ($\widehat{\beta}_{15}$)	-6	24.82
Year 4 General Assistance Effect ($\widehat{\beta}_{16}$)	3	26.53

(2020), as follows:

$$\begin{aligned} MVPF_{\text{Job Start}} = f(\widehat{\beta}) &= \frac{\sum_{i=1}^{4} \widehat{\beta}_i \times 0.993 + \sum_{i=5}^{16} \widehat{\beta}_i + 606.13}{4548 + \sum_{i=5}^{16} \widehat{\beta}_i} \\ &= 0.20 \end{aligned}$$

SDP.1 is maximized when the asymptotic correlation across all coefficients is set to +1. Below, we report the estimated weights associated with the corresponding correlation coefficient ρ_{ij} in the asymptotic variance expression of $f(\hat{\boldsymbol{\beta}})$ derived in Equation 3.2. For instance, the value corresponding to $\widehat{\beta}_1$ (Column 1) and $\widehat{\beta}_2$ (Row 2) is equal to $\hat{\sigma}_1 \hat{\sigma}_2 \frac{\partial f(\hat{\boldsymbol{\beta}})}{\partial \beta_1} \frac{\partial f(\hat{\boldsymbol{\beta}})}{\partial \beta_2}$:

	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$	$\widehat{\beta}_4$	$\widehat{\beta}_5$	$\widehat{\beta}_6$	$\widehat{\beta}_7$	$\widehat{\beta}_8$	$\widehat{\beta}_9$	$\widehat{\beta}_{10}$	$\widehat{\beta}_{11}$	$\widehat{\beta}_{12}$	$\widehat{\beta}_{13}$	$\widehat{\beta}_{14}$	$\widehat{\beta}_{15}$	$\widehat{\beta}$
$\widehat{\beta}_1$																
$\widehat{\beta}_2$	0.00146															
$\widehat{\beta}_3$	0.00181	0.00250														
$\widehat{\beta}_4$	0.00187	0.00258	0.00319													
$\widehat{\beta}_5$	0.00031	0.00042	0.00052	0.00054												
$\widehat{\beta}_6$	0.00036	0.00049	0.00061	0.00063	0.00010											
$\widehat{\beta}_7$	0.00049	0.00067	0.00083	0.00086	0.00014	0.00016										
$\widehat{\beta}_8$	0.00048	0.00067	0.00082	0.00085	0.00014	0.00016	0.00022									
$\widehat{\beta}_9$	0.00020	0.00028	0.00035	0.00036	0.00006	0.00007	0.00009	0.00009								
$\widehat{\beta}_{10}$	0.00020	0.00027	0.00034	0.00035	0.00006	0.00007	0.00009	0.00009	0.00004							
$\widehat{\beta}_{11}$	0.00023	0.00032	0.00040	0.00041	0.00007	0.00008	0.00011	0.00011	0.00004	0.00004						
$\widehat{\beta}_{12}$	0.00026	0.00035	0.00044	0.00045	0.00007	0.00009	0.00012	0.00012	0.00005	0.00005	0.00006					
$\widehat{\beta}_{13}$	0.00013	0.00018	0.00023	0.00024	0.00004	0.00005	0.00006	0.00006	0.00003	0.00002	0.00003	0.00003				
$\widehat{\beta}_{14}$	0.00009	0.00012	0.00015	0.00015	0.00002	0.00003	0.00004	0.00004	0.00002	0.00002	0.00002	0.00002	0.00001			
$\widehat{\beta}_{15}$	0.00014	0.00019	0.00024	0.00025	0.00004	0.00005	0.00006	0.00006	0.00003	0.00003	0.00003	0.00003	0.00002	0.00001		
$\widehat{\beta}_{16}$	0.00015	0.00021	0.00026	0.00027	0.00004	0.00005	0.00007	0.00007	0.00003	0.00003	0.00003	0.00004	0.00002	0.00001	0.00002	

In Figure 2, we assume that the earnings effect in each year is negatively correlated with the AFDC, Food Stamps, and General Assistance effects in that year.

D.6 Medicare Part D

The estimates used to construct the MVPF for Medicare Part D are drawn from Wettstein (2020). The effect on labor force participation is estimated using the same specification as in Column 1, Table 1 of Wettstein (2020). The effect on income is estimated using the same specification as in in Column 1, Table 3 of Wettstein (2020). The semi-elasticity of demand for insurance is estimated using the procedure described in Appendix Section D of Wettstein (2020).

Table D.6: MVPF Calculation for Introduction of Medicare Part D

	(1)	(2)
	Estimate	SE
Effect on Labor Force Participation ($\widehat{\beta}_1$)	-0.10	0.03
Effect on Income ($\widehat{\beta}_2$)	-6665.40	1986.92
Semi-Elasticity of Demand for Insurance ($\widehat{\beta}_3$)	0.14	0.02

We replicate the construction of the MVPF for Introduction of Medicare Part D from Wettstein (2020), as follows:

$$\begin{aligned} MVPF_{\text{Medicare Part D}} = f(\widehat{\beta}) &= \frac{0.65 \times \frac{\widehat{\beta}_1 \times -100}{25000} \times \frac{6126}{0.4}}{(0.65 + 0.65 \times \frac{\widehat{\beta}_3}{0.887} - \widehat{\beta}_1 - 0.28 \times \frac{\widehat{\beta}_2}{1588})/0.65} \\ &= 1.37 \end{aligned}$$

The reason our MVPF estimate of the introduction of Medicare Part D departs from the one in Wettstein (2020) is that we use their unconditional estimates on labor force participation and income for simplicity. The correlation matrix that maximizes **SDP.1** is:

	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$
$\widehat{\beta}_1$			
$\widehat{\beta}_2$	-1		
$\widehat{\beta}_3$	1	-1	

Below, we report the estimated weights associated with the corresponding correlation coefficient ρ_{ij} in the asymptotic variance expression of $f(\hat{\boldsymbol{\beta}})$ derived in Equation 3.2. For instance, the value corresponding to $\widehat{\beta}_1$ (Column 1) and $\widehat{\beta}_2$ (Row 2) is equal to $\hat{\sigma}_1\hat{\sigma}_2\frac{\partial f(\hat{\boldsymbol{\beta}})}{\partial\beta_1}\frac{\partial f(\hat{\boldsymbol{\beta}})}{\partial\beta_2}$:

	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$
$\widehat{\beta}_1$			
$\widehat{\beta}_2$	-0.08561		
$\widehat{\beta}_3$	0.00322	-0.002	

D.7 Foster Care

The estimates used to construct the MVPF for Foster Care are drawn from Column 3, Table 8 in Baron and Gross (2025). We replicate the construction of the MVPF for Foster Care

Table D.7: MVPF Calculation for Foster Care

	(1)	(2)
	Estimate	SE
Society's Willingness to Pay ($\widehat{\beta}_1$)	83854	29715
Cost Savings to the Government ($\widehat{\beta}_2$)	12188	6212

from Baron and Gross (2025), as follows:

$$MVPF_{\text{Foster Care}} = f(\widehat{\beta}) = \frac{\widehat{\beta}_1}{49920 - \widehat{\beta}_2}$$
$$= 2.22$$

SDP.1 is maximized when the asymptotic correlation between $\widehat{\beta}_1$ and $\widehat{\beta}_2$ is equal to 1, and the estimated weight associated with the correlation coefficient ρ_{12} in the asymptotic variance expression of $f(\hat{\boldsymbol{\beta}})$, denoted by $\hat{\sigma}_1\hat{\sigma}_2\frac{\partial f(\hat{\boldsymbol{\beta}})}{\partial\beta_1}\frac{\partial f(\hat{\boldsymbol{\beta}})}{\partial\beta_2}$, is equal to 0.28814.

D.8 UI Extension

The estimates used to construct the MVPF for UI Extension are drawn from Huang and Yang (2021). We replicate the construction of the MVPF for extension of Unemployment

Table D.8: MVPF Calculation for UI Extension

	(1)	(2)
	Estimate	SE
Effect on Transfers from UI ($\widehat{\beta}_1$)	0.038	0.009
Effect on Transfers from Re-employment bonus($\widehat{\beta}_2$)	0.019	0.011
Effect on Benefit Duration ($\widehat{\beta}_3$)	56.91	1.96
Effect on Unemployment Duration ($\widehat{\beta}_4$)	36.90	6.90

Insurance from Huang and Yang (2021), as follows:

$$
\begin{aligned}
MVPF_{\text{UI Extension}} = f(\widehat{\beta}) &= \frac{0.77 \times \frac{\widehat{\beta}_1 + (\widehat{\beta}_2/2)}{\widehat{\beta}_2} + 0.23}{1 + (1/72.9) \times (\widehat{\beta}_3 - 55.8 - 0.5 \times (\widehat{\beta}_3 - 55.8) + 0.12 \times \widehat{\beta}_4)} \\
&= 2.02
\end{aligned}
$$

The correlation matrix that maximizes SDP.1 is:

	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$	$\widehat{\beta}_4$
$\widehat{\beta}_1$				
$\widehat{\beta}_2$	-1			
$\widehat{\beta}_3$	-1	1		
$\widehat{\beta}_4$	-1	1	1	

Below, we report the estimated weights associated with the corresponding correlation coefficient ρ_{ij} in the asymptotic variance expression of $f(\hat{\boldsymbol{\beta}})$ derived in Equation 3.2. For instance, the value corresponding to $\widehat{\beta}_1$ (Column 1) and $\widehat{\beta}_2$ (Row 2) is equal to $\hat{\sigma}_1\hat{\sigma}_2\frac{\partial f(\hat{\boldsymbol{\beta}})}{\partial\beta_1}\frac{\partial f(\hat{\boldsymbol{\beta}})}{\partial\beta_2}$:

	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$	$\widehat{\beta}_4$
$\widehat{\beta}_1$				
$\widehat{\beta}_2$	-0.28491			
$\widehat{\beta}_3$	-0.00867	0.02118		
$\widehat{\beta}_4$	-0.00732	0.01790	0.00054	