On the minimum strength of (unobserved) covariates to overturn an insignificant result

Danielle Tsao, Ronan Perry and Carlos Cinelli

Abstract. We study conditions under which the addition of variables to a regression equation can turn a previously statistically insignificant result into a significant one. Specifically, we characterize the minimum strength of association required for these variables—both with the dependent and independent variables, or with the dependent variable alone—to elevate the observed t-statistic above a specified significance threshold. Interestingly, we show that it is considerably difficult to overturn a statistically insignificant result solely by reducing the standard error. Instead, included variables must also alter the point estimate to achieve such reversals in practice. Our results can be used for sensitivity analysis and for bounding the extent of p-hacking, and may also offer algebraic explanations for patterns of reversals seen in empirical research, such as those documented by Lenz and Sahn (2021).

Key words and phrases: ordinary least squares, linear regression, omitted variable bias, robustness values, p-hacking.

1. INTRODUCTION

Applied researchers are often confronted with unexpected statistically insignificant estimates of linear regression coefficients. This situation may lead to the addition of variables to the regression equation with the intent to reduce standard errors or to account for factors that could be masking the target relationship of interest (Cinelli, Forney and Pearl, 2022). If statistical significance remains elusive, researchers may naturally speculate whether there exist key variables that remained unmeasured but could have overturned statistical insignificance had they been accounted for in the analysis. As statistical significance is often a key factor for publication, such practices and concerns frequently arise in both experimental and observational studies.

Consider, for example, a randomized controlled trial (RCT) in which a researcher uses ordinary least squares (OLS) to estimate the average effect of a treatment on an outcome. Here, confounding biases do not exist by design but adjusting for covariates may still help with precision. If the initial result is not statistically significant, this may lead to the inclusion of pre-treatment covariates in the regression equation to potentially attain statistical significance. In observational studies, beyond precision gains, covariate adjustment may be an essential tool for obtaining valid estimates of the target of inference. It may help mitigate confounding biases, block indirect pathways, estimate conditional effects, or address various other methodological concerns—all of which could provide legitimate reasons for introducing covariates that reverse an initially insignificant result, or to ask whether unobserved variables could have done so.

However, despite the many valid reasons for covariate adjustment, applied researchers often fail to adequately justify their choice of control variables. For example, in the American Journal of Political Science, Lenz and Sahn (2021) found that over 30% of articles relied on the inclusion of covariates to turn previously statistically insignificant findings into significant ones. According to Lenz and Sahn (2021), none of the articles justified this choice, nor disclosed these reversals. In fact, the practice of testing various model specifications with the intention of obtaining statistically significant results is commonly referred to as 'p-hacking' (Simonsohn, Nelson and Simmons, 2014). Extensive surveys and meta-analysis of published p-values suggest that p-hacking may be prevalent across disciplines (Brodeur et al., 2016; Vivalt, 2019).

Under what conditions can such reversals of statistical insignificance occur? Can we establish bounds on the extent of 'p-hacking'? And what observable patterns in the data should emerge when these reversals take place? In this short communication, we provide simple algebraic answers to these questions in the context of OLS. Building on recent results from Cinelli and Hazlett (2020, 2022), we first characterize the maximum change in the

University of Washington, Seattle, USA (e-mail: dltsao@uw.edu). University of Washington, Seattle, USA (e-mail: rflperry@uw.edu). University of Washington, Seattle, USA (e-mail: cinelli@uw.edu).

t-statistic that covariates with bounded strength can produce. We then derive the minimum strength of association that such covariates must have—whether with both the dependent and independent variables or with the dependent variable alone—to elevate the observed t-statistic above a given statistical significance threshold. Lastly, we provide an empirical example. These results can be applied to conduct sensitivity analyses against unobserved 'suppressors' and to bound the extent of p-hacking arising due to the choice of control variables. It may also offer algebraic explanations for patterns of significance reversals observed in empirical research.

2. PRELIMINARIES

2.1 Problem setup

Let Y be an $(n \times 1)$ vector containing the dependent variable for n observations; D be an $(n \times 1)$ independent variable of interest and $\mathbf{X}$ be an $(n \times p)$ matrix of observed covariates including a constant. Consider the regression equation

$$Y = \hat{\lambda}_r D + \mathbf{X}\hat{\beta}_r + \hat{\epsilon}_r, \tag{1}$$

where $\hat{\lambda}_r$, $\hat{\beta}_r$ are the OLS estimates of the regression coefficients of Y on D and X, and $\hat{\epsilon}_r$ is the corresponding $(n \times 1)$ vector of residuals.

Let $\widehat{\text{se}}(\hat{\lambda}_r)$ be the estimated classical (homoskedastic) standard error of $\hat{\lambda}_r$. Under the classical linear regression model, the t-statistic for testing the null hypothesis $H_0 : \lambda_r = \lambda_0$, i.e.,

$$t_r := \frac{\hat{\lambda}_r - \lambda_0}{\widehat{\text{se}}(\hat{\lambda}_r)}, \tag{2}$$

follows a t-distribution with $\text{df} := n - p - 1$ degrees of freedom. Denoting by $t^*_{\alpha,\text{df}}$ the $(1 - \alpha/2)$ quantile of this distribution, the t-statistic (2) is considered "statistically significant with significance level α" if the absolute value of t_r exceeds that of $t^*_{\alpha,\text{df}}$. Note that the t-statistic depends on the choice of λ_0. For simplicity, we use the notation t_r with the understanding that a particular λ_0 has been chosen.

Now suppose the t-statistic (2) is insignificant. Let Z be an $(n \times 1)$ vector of a (potentially unobserved) covariate whose inclusion in the regression equation we wish to assess. In contrast to the 'restricted' regression in (1), we now consider the long regression equation of Y on D after adjusting for both X and Z,

$$Y = \hat{\lambda} D + \mathbf{X}\hat{\beta} + \hat{\gamma} Z + \hat{\epsilon}. \tag{3}$$

Here, the t-statistic for testing null hypothesis $H_0 : \lambda = \lambda_0$ is

$$t := \frac{\hat{\lambda} - \lambda_0}{\widehat{\text{se}}(\hat{\lambda})} \tag{4}$$

where $\hat{\lambda}$ and $\widehat{\text{se}}(\lambda)$ have the same interpretation as before, just now with an additional adjustment for Z. We wish to quantify the properties that Z needs to have such that the t-statistic in (4) will be statistically significant.

2.2 Omitted variable bias formulas

Comparing (2) with (4), observe that the (absolute) relative change in the t-statistic can be decomposed as the product of the relative change in the bias and the relative change in the standard error:

$$\left|\frac{t}{t_r}\right| = \left|\frac{\hat{\lambda} - \lambda_0}{\hat{\lambda}_r - \lambda_0}\right| \times \left(\frac{\widehat{\text{se}}(\hat{\lambda}_r)}{\widehat{\text{se}}(\hat{\lambda})}\right). \tag{5}$$

Concretely, for Z to double the t-statistic, it must either double the absolute difference between the point estimate and λ_0, halve the standard errors, or achieve some combination of both.

To characterize these changes in terms of how much residual variation Z explains of D and Y, we refer to the following result from Cinelli and Hazlett (2020).

THEOREM 1 (Cinelli and Hazlett, 2020). *Let $R^2_{Y\sim Z|D\mathbf{X}}$ denote the sample partial R^2 of Y with Z after adjusting for D and $\mathbf{X}$, and let $R^2_{D\sim Z|\mathbf{X}} < 1$ denote the sample partial R^2 of D with Z after adjusting for $\mathbf{X}$. Then,*

$$|\hat{\lambda}_r - \hat{\lambda}| = \underbrace{\sqrt{\frac{R^2_{Y\sim Z|D\mathbf{X}} R^2_{D\sim Z|\mathbf{X}}}{1 - R^2_{D\sim Z|\mathbf{X}}}}}_{BF} \times \widehat{se}(\hat{\lambda}_r) \times \sqrt{df} \tag{6}$$

and

$$\widehat{se}(\hat{\lambda}) = \underbrace{\sqrt{\frac{1 - R^2_{Y\sim Z|D\mathbf{X}}}{1 - R^2_{D\sim Z|\mathbf{X}}}}}_{SEF} \times \widehat{se}(\hat{\lambda}_r) \times \sqrt{\frac{df}{df - 1}}. \tag{7}$$

To aid interpretation, we call the terms BF in (6) and SEF in (7) the "bias factor" and the "standard error factor" of Z, respectively.

We can use Theorem 1 to write the absolute value of the t-statistic (4) as a function of $R^2_{Y\sim Z|D\mathbf{X}}$ and $R^2_{D\sim Z|\mathbf{X}}$, i.e.,

$$t(R^2_{Y\sim Z|D\mathbf{X}}, R^2_{D\sim Z|\mathbf{X}}) = \frac{|(\hat{\lambda}_r - \lambda_0) \pm \text{BF} \times \widehat{\text{se}}(\hat{\lambda}_r) \times \sqrt{\text{df}}|}{\widehat{\text{se}}(\hat{\lambda}_r) \times \text{SEF} \times \sqrt{\frac{\text{df}}{\text{df}-1}}},$$

where the sign of the bias term, denoted by $\pm$, depends on whether $\hat{\lambda}_r > \hat{\lambda}$ or vice-versa. This re-formulation allows us to assess how Z affects inferences for any postulated pair of partial R^2 values $\{R^2_{Y\sim Z|D\mathbf{X}}, R^2_{D\sim Z|\mathbf{X}}\}$, and it will help us determine the conditions under which the addition of Z turns a previously statistically insignificant result into a significant one.

To set the stage for upcoming results, we note an immediate but important corollary of Theorem 1: for a fixed observed t-statistic and fixed strength of Z, the impact that Z has on the relative bias depends on the sample size, whereas the impact it has on the relative change in standard errors does not. The relative change in the bias is given by

$$\frac{\hat{\lambda}-\lambda_0}{\hat{\lambda}_r-\lambda_0} = 1 \pm \frac{\text{BF}}{t_r} \times \sqrt{\text{df}}.$$

Notice that for fixed t_r and fixed $\{R^2_{Y\sim Z|D\mathbf{X}}, R^2_{D\sim Z|\mathbf{X}}\}$ (which sets BF), larger sample sizes yield larger relative changes in the distance of the estimate from the null hypothesis. Conversely, the relative change in the standard error is given by

$$\widehat{\text{se}}(\hat{\lambda}_r)/\widehat{\text{se}}(\hat{\lambda}) = \frac{1}{\text{SEF}} \times \sqrt{\frac{\text{df}-1}{\text{df}}} \approx \frac{1}{\text{SEF}} \tag{8}$$

and is thus unaffected by the sample size. As an example, halving standard errors is equally challenging in a sample of 100 as in a sample of 1,000,000; in contrast, doubling point estimates becomes much easier as the sample size grows. This distinction will become clearer as we continue with our analysis.

3. RESULTS

In this section we present two main results. First, given upper-bounds on $R^2_{Y\sim Z|D\mathbf{X}}$ and $R^2_{D\sim Z|\mathbf{X}}$, we derive the 'maximum adjusted t-statistic' which quantifies the maximum possible value that $t(R^2_{Y\sim Z|D\mathbf{X}}, R^2_{D\sim Z|\mathbf{X}})$ can attain after including Z in the regression equation. Then we solve for the minimum upper bound on $\{R^2_{Y\sim Z|D\mathbf{X}}, R^2_{D\sim Z|\mathbf{X}}\}$, hereby referred to as the "strength of Z", such that it guarantees that the maximum adjusted t-statistic exceeds the desired significance threshold.

In what follows, it is useful to define the quantities

$$f_r := |t_r|/\sqrt{\text{df}} \quad \text{and} \quad f^*_{\alpha,\text{df}} := t^*_{\alpha,\text{df}}/\sqrt{\text{df}},$$

which normalize the observed t-statistic and the critical threshold by the degrees of freedom. These definitions greatly simplify formulas and derivations.

3.1 On the maximum adjusted t-statistic

We start by defining the maximum value that the t-statistic (4) could attain given Z with bounded strength.

DEFINITION 1 (Maximum adjusted t-statistic). For a fixed null hypothesis $H_0 : \lambda = \lambda_0$, significance level α, and upper bounds on $R^2_{Y\sim Z|D\mathbf{X}}$ and $R^2_{D\sim Z|\mathbf{X}}$, denoted by $\boldsymbol{R}^2 = \{R^{\max}_Y, R^{\max}_D\}$, we define the maximum adjusted t-statistic as

$$t^{\max}_{\boldsymbol{R}^2} := \max_{R^2_{Y\sim Z|D\mathbf{X}}, R^2_{D\sim Z|\mathbf{X}}} t(R^2_{Y\sim Z|D\mathbf{X}}, R^2_{D\sim Z|\mathbf{X}})$$

$$\text{s.t. } R^2_{Y\sim Z|D\mathbf{X}} \le R^{\max}_Y, R^2_{D\sim Z|\mathbf{X}} \le R^{\max}_D.$$

The solution to the above problem has a simple closed-form characterization.

THEOREM 2 (Closed-form solution to $t^{\max}_{\boldsymbol{R^2}}$). *Let $R^{\max}_Y < 1$. Then,*

$$t^{\max}_{\boldsymbol{R^2}} = \frac{f_r\sqrt{1-R^{*2}_{D\sim Z|\mathbf{X}}} + \sqrt{R^{*2}_{Y\sim Z|D\mathbf{X}} R^{*2}_{D\sim Z|\mathbf{X}}}}{\sqrt{(1-R^{*2}_{Y\sim Z|D\mathbf{X}})/(df-1)}}$$

where

$$\{R^{*2}_{Y\sim Z|D\mathbf{X}}, R^{*2}_{D\sim Z|\mathbf{X}}\} = \{R^{\max}_Y, R^{\max}_D\}$$

if $f_r^2 < R^{\max}_Y(1-R^{\max}_D)/R^{\max}_D$ and

$$\{R^{*2}_{Y\sim Z|D\mathbf{X}}, R^{*2}_{D\sim Z|\mathbf{X}}\} = \left\{R^{\max}_Y, \frac{R^{\max}_Y}{f_r^2+R^{\max}_Y}\right\}$$

otherwise.

If $t^{\max}_{\boldsymbol{R}^2} < t^*_{\alpha,\text{df}-1}$, then we can be assured that no Z with the specified maximum strength would be able to overturn an insignificant result. On the other hand, if $t^{\max}_{\boldsymbol{R}^2} > t^*_{\alpha,\text{df}-1}$, we know that there exists at least one Z with strength no greater than $\boldsymbol{R}^2$ that is capable of bringing the t-statistic above the specified threshold.

REMARK 1. Note that the optimal value of $R^2_{Y\sim Z|D\mathbf{X}}$ always reaches the upper bound $R^{\max}_Y$, while $R^2_{D\sim Z|\mathbf{X}}$ may either reach its upper bound $R^{\max}_D$ or result in the interior point solution $\frac{R^{\max}_Y}{f_r^2+R^{\max}_Y}$.

REMARK 2. It is always necessary to constrain the strength of Z with respect to Y in order to obtain a finite solution for $t^{\max}_{\boldsymbol{R}^2}$. If $R^{\max}_Y = 1$, then as $R^2_{Y\sim Z|D\mathbf{X}}$ approaches one, the standard error approaches zero and the t-statistic grows without bounds.

REMARK 3. In contrast, it is possible to leave $R^2_{D\sim Z|\mathbf{X}}$ unconstrained. Increasing $R^2_{D\sim Z|\mathbf{X}}$ has two counterbalancing effects on the t-statistic. On one hand, it can change the point estimate, as described by (6), in a direction that is favorable for rejecting $H_0 : \lambda = \lambda_0$. On the other hand, it also increases the standard error due to the variance inflation factor in (7) (i.e. the denominator of the SEF), which eventually counter-balances and then exceeds the benefit of the change in estimate. Thus, setting $R^{\max}_D = 1$ will always result in an interior point solution for $R^2_{D\sim Z|\mathbf{X}}$.

Of course, there naturally could be multiple latent variables instead of a single one, and so one might wonder about the case when Z takes the form of a matrix rather than a vector. The following theorem demonstrates that it is sufficient to consider a single unmeasured latent variable, up to a correction in the degrees of freedom.

THEOREM 3 ($t^{\max}_{\boldsymbol{R}^2}$ for matrix $\mathbf{Z}$). *Let $\mathbf{Z}$ be an $(n \times m)$ matrix of covariates. Then the solution to $t^{\max}_{\boldsymbol{R}^2}$ is the same as that of Theorem 2, save for the adjustment in the degrees of freedom, which now is $df - m$.*

In what follows, for simplicity we keep Z as a vector with the understanding that all results still hold for matrix $\mathbf{Z}$. But before moving forward, there is an interesting corollary of the previous result. It places limits on the extent of p-hacking given observed covariates $\mathbf{X}$.

COROLLARY 1 (Upper bound on p-hacking). *For observed covariates $\mathbf{X}$, let $R^2_{Y\sim\mathbf{X}|D}, R^2_{D\sim\mathbf{X}}$ denote the strengths of the associations of $\mathbf{X}$ with Y and D, respectively. Let $t_{Y\sim D|\mathbf{X}_S}$ denote the t-statistic for the coefficient of the regression of D on Y when adjusting for the subset of covariates $\mathbf{X}_S$ (consisting of a subset of the columns of $\mathbf{X}$). Then for any $\mathbf{X}_S$,*

$$t_{Y\sim D|\mathbf{X}_S} \leq t^{\max}_{\boldsymbol{R}^2},$$

where $t^{\max}_{\boldsymbol{R}^2}$ is the solution of Theorem 2 using $t_r = t_{Y\sim D}$ and $\boldsymbol{R}^2 = \{R^2_{Y\sim\mathbf{X}|D}, R^2_{D\sim\mathbf{X}}\}$.

When the number of covariates is small, it is feasible to run all possible regressions to identify the exact maximum t-statistic across all specifications. However, when the number of covariates is large, this exhaustive approach becomes impractical. For example, with $p = 40$, there are 2^{40} (approximately 1 trillion) possible specifications. In such a case, $t^{\max}_{\boldsymbol{R}^2}$ offers a simple upper bound on the maximum extent of p-hacking without the need to run all 1 trillion regressions.

EXAMPLE 1. Let $p = 40$, $\text{df} = 100$ and the t-statistic of the regression of Y on D be equal to 1. Then, if $\{R^2_{Y\sim\mathbf{X}|D}, R^2_{D\sim\mathbf{X}}\} = \{0.08, 0.08\}$, Corollary 1 assures us that none of the 1 trillion specifications can yield a t-statistic greater than 1.83.

REMARK 4. Note that $t^{\max}_{\boldsymbol{R}^2}$ is achievable when the only constraint on the variables to be included is their maximum explanatory power. For any given set of observed covariates, $t^{\max}_{\boldsymbol{R}^2}$ is a potentially loose upper bound. It is possible to tighten this bound by applying the corollary iteratively within subsets of regressions. Obtaining tight bounds without running all of the 2^m possible regressions remains an open problem.

3.2 On the minimal strength of Z to reverse statistical insignificance

Equipped with the notion of the maximum adjusted t-statistic, we can now characterize the minimum strength of Z necessary to obtain a statistically significant result. Following the convention of Cinelli and Hazlett (2020), we call our metrics "robustness values" for insignificance. They quantify how "robust" an insignificant result is to the inclusion of covariates in the regression equation.

3.2.1 Extreme robustness value for insignificance. As highlighted in Remark 2, the parameter $R^2_{Y\sim Z|D\mathbf{X}}$ is essential for assessing the potential of Z to bring about a significant result, as it always needs to be bounded. Thus we begin by characterizing the minimal strength of association of Z with Y *alone* in order to achieve significance.

DEFINITION 2 (Extreme Robustness Value for Insignificance). For fixed $R^{\max}_D \in [0,1]$, the extreme robustness value for insignificance, $\text{XRVI}^{R^{\max}_D}_\alpha$, is the minimum upper bound on $R^2_{Y\sim Z|D\mathbf{X}}$ such that $t^{\max}_{\boldsymbol{R}^2}$ is large enough to reject null hypothesis $H_0 : \lambda = \lambda_0$ at specified significance level α, i.e.,

$$\text{XRVI}^{R^{\max}_D}_\alpha := \min\{\text{XRVI} : t^{\max}_{\text{XRVI}, R^{\max}_D} \geq t^*_{\alpha,\text{df}-1}\}.$$

For a fixed bound on $R^2_{D\sim Z|\mathbf{X}}$, the $\text{XRVI}^{R^{\max}_D}_\alpha$ describes how robust an insignificant result is in terms of the minimum explanatory power that Z needs to have with Y in order to overturn it. Theorem 7 in the appendix provides an analytical expression for $\text{XRVI}^{R^{\max}_D}_\alpha$ given arbitrary $R^{\max}_D \in [0,1]$. Here we focus on two important cases: $R^{\max}_D = 0$ and $R^{\max}_D = 1$.

Starting with $R^{\max}_D = 0$, we first consider the scenario where the point estimate remains unchanged, and any increase in the t-statistic occurs solely due to a reduction in the standard error. That is, XRVI^0_α quantifies how much variation a control variable Z that is uncorrelated with D must explain of the dependent variable Y in order to overturn a previously insignificant result. This turns out to have a remarkably simple and insightful characterization.

THEOREM 4 (Closed-form expression for XRVI^0_α). *Let $f_r > 0$, then the analytical solution for $XRVI^0_\alpha$ is*

$$XRVI^0_\alpha = \begin{cases} 0, & \text{if } f^*_{\alpha,df-1} < f_r, \\ 1 - \left(\dfrac{f_r}{f^*_{\alpha,df-1}}\right)^2, & \text{otherwise.} \end{cases}$$

If $f_r = 0$, there is no value of $R^2_{Y\sim Z|D\mathbf{X}}$ capable of overturning an insignificant result.

REMARK 5. It is useful to understand how XRVI^0_α changes as the sample size grows, when keeping the observed t-statistic and the significance level fixed.

$$\text{XRVI}^0_\alpha \approx 1 - \left(\frac{t_r}{t^*_{\alpha,\text{df}-1}}\right)^2 \xrightarrow[\infty]{\text{df}} 1 - \left(\frac{t_r}{z^*_\alpha}\right)^2,$$

where here $\xrightarrow[\infty]{\text{df}}$ denotes the limit as df goes to infinity and z^*_α denotes the $(1-\alpha/2)$ quantile of the standard normal distribution. In other words, when considering only a reduction in the standard error, the amount of residual variation that Z must explain of Y in order to overturn an insignificant result depends solely on the ratio of the observed t-statistic to the critical threshold. Apart from changes in the critical threshold due to degrees of freedom— which eventually converges to z^*_α—this value remains constant regardless of sample size.

EXAMPLE 2. Consider testing the null hypothesis of zero effect with an observed t-statistic of 1 and 100 degrees of freedom. The percentage of residual variation of Y that Z needs to explain in order to bring a t-statistic of 1 to the critical threshold of 2, *only through a reduction in the standard error*, is

$$\text{XRVI}^0_\alpha \approx 1-\left(\frac{1}{2}\right)^2 = 1-(1/4) = 3/4 = 75\%.$$

That is, if we are considering a reduction in the standard error alone, Z needs to explain at least 75% of the variation of Y in order to elevate the observed the t-statistic to 2. Notably, this number is (virtually) the same across sample sizes, be it $\text{df} = 100$, $\text{df} = 1,000$ or $\text{df} = 1,000,000$. As variables that explain 75% of the variation of Y are rare in most settings, this simple fact suggests that it should also be rare to see a reversal of significance driven by gains in precision when $t_r = 1$. We return to this point in the discussion.

The previous result describes how to achieve statistical significance via precision gains. We now move to the case with $R^{\max}_D = 1$. As noted in Remark 3, this is the scenario in which we *impose no constraints* on the strength of association between Z and D. In this sense, XRVI^1_α computes the *bare minimum* amount of variation that Z needs to explain of Y in order to reverse an insignificant result. Any variable that does not explain at least $(100\times\text{XRVI}^1_\alpha)\%$ of the variation of Y is logically incapable of making the t-statistic significant.

THEOREM 5 (Closed-form expression for XRVI^1_α). *The analytical solution for $XRVI^1_\alpha$ is:*

$$XRVI^1_\alpha = \begin{cases} 0, & \text{if } f^*_{\alpha,df-1} < f_r, \\ \dfrac{f^{*2}_{\alpha,df-1} - f^2_r}{1+f^{*2}_{\alpha,df-1}}, & \text{otherwise.} \end{cases}$$

REMARK 6. Contrary to the previous case, we observe the following behaviour as the sample size grows,

$$\text{XRVI}^1_\alpha \approx \left(\frac{t^{*2}_{\alpha,\text{df}-1} - t^2_r}{\text{df} + t^{*2}_{\alpha,\text{df}-1}}\right) \xrightarrow[\infty]{\text{df}} 0.$$

Therefore, if we allow Z to change point estimates, then for a fixed observed t-statistic, the minimal strength of Z with Y to bring about a reversal tends to zero as the sample size grows to infinity.

EXAMPLE 3. Consider again testing the null hypothesis of zero effect with an observed t-statistic of 1 and 100 degrees of freedom. If we allow Z to be arbitrarily associated with D, it needs only to explain 2.9% of the residual variation of Y in order to bring the t-statistic to 2:

$$\text{XRVI}^1_\alpha \approx \frac{2^2-1^2}{100+2^2} = \frac{3}{104} = 2.9\%.$$

Also note that any Z that explains less than 2.9% of the variation of Y is logically incapable of bringing about such change. Corroborating our previous analysis, the Z that achieves this must do so via an increase in point estimate, and not via a decrease in standard errors. As per Theorem 2, the optimal value of the association with D is $R^2_{D\sim Z|\mathbf{X}} \approx 74\%$. Notice that $\text{SEF} \approx 1.94$, meaning that the inclusion of Z almost *doubles* the standard error, instead of reducing it. This, however, is compensated by the fact that Z increases the point estimate by a factor of $1+\text{BF}\times\sqrt{\text{df}} = 3.88$, thus doubling the t-statistic despite the loss in precision.

EXAMPLE 4. For the same observed t-statistic of 1, consider a sample size that is an order of magnitude larger, say, $\text{df} = 1,000$. The minimum residual variation that Z needs to explain of Y then reduces to 0.29%:

$$\text{XRVI}^1_\alpha \approx \frac{2^2-1^2}{1000+2^2} = \frac{3}{1004} = 0.29\%.$$

As per Theorem 2, this Z has an association with D of $R^2_{D\sim Z|\mathbf{X}} \approx 75\%$. Here, we have the same situation as before: the standard error doubles while the point estimate increases by a factor of four, thus doubling the t-statistic.

3.2.2 Robustness value for insignificance. While in the previous section we investigated the minimal bound on $R^2_{Y\sim Z|D\mathbf{X}}$ alone in order to revert an insignificant result, here we investigate the minimal bound on both $R^2_{Y\sim Z|D\mathbf{X}}$ and $R^2_{D\sim Z|\mathbf{X}}$ simultaneously.

DEFINITION 3 (Robustness Value for Insignificance). The robustness value for insignificance, RVI_α, is the minimum upper bound on both $R^2_{Y\sim Z|D\mathbf{X}}$ and $R^2_{D\sim Z|\mathbf{X}}$ such that $t^{\max}_{\boldsymbol{R}^2}$ is large enough to reject the null hypothesis $H_0: \lambda = \lambda_0$ at specified significance level α. That is,

$$\text{RVI}_\alpha := \min\{\text{RVI} : t^{\max}_{\text{RVI,RVI}} \geq t^*_{\alpha,\text{df}-1}\}.$$

Note that RVI_α provides a convenient summary of the minimum strength of association that Z needs to have, *jointly* with D and Y, in order to bring about a statistically significant result. Any Z that has both partial R^2 values no stronger than RVI_α cannot reverse a statistically insignificant finding. On the other hand, we can always find a Z with both partial R^2 values at least as strong as RVI_α that does so. The solution of this problem is given in the following result.

THEOREM 6 (Closed-form expression for RVI_α). *The analytical solution for RVI_α is*

$$RVI_\alpha = \begin{cases} 0, & \text{if } f^*_{\alpha,df-1} < f_r, \\ \frac{1}{2}\left(\sqrt{f_\Delta^4+4f_\Delta^2}-f_\Delta^2\right), & \text{if } f_r < f^*_{\alpha,df-1} < f_r^{-1}, \\ XRVI^1_\alpha, & \text{otherwise} \end{cases}$$

where $f_\Delta := f^*_{\alpha,df-1} - f_r$.

REMARK 7. The first case in Theorem 6 occurs when the t-statistic for $H_0 : \lambda = \lambda_0$ is already statistically significant, even when losing one degree of freedom. The second case occurs when both constraints on $R^2_{D\sim Z|\mathbf{X}}$ and $R^2_{Y\sim Z|D\mathbf{X}}$ are binding. The third case is the interior point solution, as defined in Theorem 5, where only the constraint on $R^2_{Y\sim Z|D\mathbf{X}}$ is binding.

Notice that in the second solution of RV_α, we have $R^2_{D\sim Z|\mathbf{X}} = R^2_{Y\sim Z|D\mathbf{X}}$; thus, SEF $= 1$ and standard errors remain unchanged. Therefore, RV_α represents the minimal strength of Z needed to achieve statistical significance via a *change in the point estimate alone*, usefully complementing XRVI^0_α, which quantifies the minimal strength of Z needed solely through a reduction in standard errors.

REMARK 8. We recover the XRVI as the solution to RVI if and only if the conditions (1) $f^*_{\alpha,\mathrm{df}-1} > f_r^{-1}$ and (2) $f^*_{\alpha,\mathrm{df}-1} > f_r$ both hold. This rarely occurs. To see this more clearly, note that condition (1) simplifies to $\sqrt{\mathrm{df}-1}\sqrt{\mathrm{df}} < t_r \times t^*_{\alpha,\mathrm{df}-1}$, or, approximately,

$$\mathrm{df} \lessapprox t_r \times t^*_{\alpha,\mathrm{df}-1}$$

which only occurs when there are few degrees of freedom, for typical critical thresholds (e.g. 1.96).

REMARK 9. As with XRVI^1_α, for fixed t_r and significance level α, we observe the same behaviour for RVI_α as the sample size grows,

$$\mathrm{RVI}_\alpha \approx \frac{1}{2}\left(\sqrt{\frac{t_\Delta^4}{\mathrm{df}^2} + 4\frac{t_\Delta^2}{\mathrm{df}}} - \frac{t_\Delta^2}{\mathrm{df}}\right) \xrightarrow[\infty]{\mathrm{df}} 0,$$

where $t_\Delta = t^*_{\alpha,\mathrm{df}-1} - t_r$. Therefore, the larger the sample size, any change in the point estimate will eventually be sufficiently strong to bring about statistical significance.

REMARK 10. The statistics we introduced here obey the following ordering,

$$\mathrm{XRVI}^1_\alpha \leq \mathrm{RVI}_\alpha \leq \mathrm{XRVI}^0_\alpha.$$

This follows directly from their definitions, as each case represents a constrained minimization problem and the constraint becomes stricter as we move from $R^{\max}_D = 1$ to $R^{\max}_D = 0$. Moreover, $\mathrm{XRVI}^{\mathrm{RVI}_\alpha}_\alpha = \mathrm{RVI}_\alpha$.

EXAMPLE 5. Continuing with the case where the t-statistic is 1, we obtain $\mathrm{RVI}_\alpha \approx 9.5\%$ when $\mathrm{df} = 100$ and $\mathrm{RVI}_\alpha \approx 3\%$ when $\mathrm{df} = 1,000$. In both cases, the inflation of the t-statistic by a Z that attains the optimal strength is driven solely by changes in the point estimate.

4. EMPIRICAL EXAMPLE

We demonstrate the use of our metrics in an empirical example that estimates the effect of a vote-by-mail policy in various outcomes (Amlani and Collitt, 2022). This work includes an analysis for the effect of a US county's vote-by-mail (VBM) policy on the turnout and the Republican presidential vote share (dependent variable Y) in the 2020 election. There are 5 treatment conditions concerning the VBM policy change from 2016 to 2020, of which the authors are specifically interested in the condition: no-excuse-needed (in 2016) to ballots-sent-in (in 2020), which we will refer to as *condition-1*. The authors fit a differences-in-differences model using OLS, adjusting for various covariates including an indicator for battleground states and the median age and the median income of residents in the county. Notably, the interaction term for *condition-1* $\times$ *year* (independent variable D) is not statistically significant at the 5% level: the t-statistic is $t_r = 0.12$ with $4,307$ degrees of freedom.

4.1 Robustness to unobserved suppressors

The authors were concerned that the lack of significance for the coefficient of interest could have been due to suppression effects of unobserved variables. To address this, they use the formulas of Theorem 1 to examine whether different hypothetical values for the strength of Z yield a statistically significant t-statistic. Here we complement their analysis by providing the three proposed metrics, XRVI^1, RV and XRV^0.

The results are displayed in Table 1. We find that any latent variable Z that explains less than 2.77% of the residual variation of both Y and D ($\mathrm{RVI}_\alpha = 2.77\%$) would

Estimate	Std. Error	t-statistic	XRVI^1_α	RVI_α	XRVI^0_α
0.103	0.873	0.118	0.089%	2.77%	99.6%

Note: df = 4307, $\lambda_0 = 0$, $\alpha = 0.05$.

TABLE 1
Robustness values for insignificance for the vote-by-mail policy study.

not be sufficiently strong to make the estimate statistically significant. Moreover, if we impose no constraints on $R^2_{D\sim Z|\mathbf{X}}$, then Z needs to explain at least 0.089% of the variation of Y in order to attain such a reversal ($\text{XRVI}^1_\alpha = 0.089\%$) . Finally, our analysis shows that a reversal of significance solely due to gains in precision is virtually impossible: a Z orthogonal to D would need to explain a remarkable 99.6% of the variation of Y in order to overturn the insignificant result ($\text{XRVI}^0_\alpha = 99.6\%$). To put these statistics in context, a latent variable with the same strength of association with D and Y as that of the battleground state indicator would only explain 0.6% of the residual variation in Y and 0.27% of the residual variation in D. Since both of these numbers are below the RVI_α value of 2.77%, we can immediately conclude that adjusting for a latent variable Z of similar strength to this observed covariate would not be sufficient to overturn the insignificant result.

4.2 Robustness to subsets of controls

We now illustrate how $t^{\max}_{\boldsymbol{R}^2}$ can be used to understand whether one can easily rule out the possibility of obtaining a statistically significant t-statistic when adjusting for different subsets of control variables. The authors present three model specifications, all of which found the effect of interest to be statistically insignificant. However, could there be a specification where the results turn out to be significant? Here we consider all possible variations between their base model and an expanded model that includes 12 additional control variables. This amounts to $2^{12} = 4{,}096$ possible regressions. Applying the results of Corollary 1, we obtain $t^{\max}_{\boldsymbol{R}^2} \approx 20.7$, meaning that we cannot rule out that there exists a specification where the interaction term becomes significant. Given the relatively small number of combinations, we can actually compute the ground truth to verify—and, indeed there are 510 models that yield a statistically significant result.

5. DISCUSSION

The algebra of OLS both imposes strong limits on and reveals clear patterns in how reversals of statistical insignificance occur. The first lesson that emerges from our analysis is that reversals of low t-statistics are unlikely to occur through reductions in standard errors alone. As shown in the first row of Table 2, even with a t-statistic of 1.75, one would need to explain at least 20% of the residual variation in Y to achieve statistical significance at the 5% level solely through a reduction in standard errors. Elevating a t-statistic of .5 to statistical significance requires explaining a remarkable 93% of the residual variation of Y. Such large strengths of association with the response variable are not typically common in many empirical applications.

t_r	0.25	0.50	.75	1.00	1.25	1.50	1.75
$\text{XRVI}^0_{\alpha=0.05}$	0.98	0.93	0.85	0.74	0.59	0.41	0.20
$\text{XRVI}^{q.95}_{\alpha=0.05}$	0.41	0.32	0.24	0.16	0.10	0.05	0.01

TABLE 2
Approximate values of XRVI for various values of t_r.

A second consequence of our findings is that, in RCTs, it should be difficult to observe reversals of low t-statistics due to covariate adjustments. Since covariates in such trials typically have zero association with the treatment by design (barring sampling errors), their inclusion is unlikely to significantly shift the point estimate. A back-of-the-envelope calculation illustrates this point: if D is randomized, then df $R^2_{D\sim Z|\mathbf{X}}$ follows an approximate chi-square distribution with one degree of freedom. Thus, letting $q_{.95}$ denote the (approximate) 95th percentile of realizations of $R^2_{D\sim Z|\mathbf{X}}$, we can calculate $\text{XRVI}^{q.95}_{0.05}$ for various values of t_r. These values are recorded in the second line of Table 2. With the exception of $t_r = 1.75$ and $t_r = 1.5$, the values of $R^2_{Y\sim Z|D\mathbf{X}}$ required to reverse statistical insignificance remain moderate to large, suggesting that such reversals should be uncommon in practice.

Finally, and perhaps counter-intuitively, even when included variables are highly predictive of the response, reversals of insignificance are still typically driven by shifts in the point estimate rather than by reductions in standard errors. To illustrate, consider the usual critical threshold of $t^*_{\alpha,\text{df}-1} \approx 2$ and any observed t-statistic below 1. Then, if $R^2_{Y\sim Z|D\mathbf{X}} \leq .5$, it is *impossible* to obtain a reversal that is not mainly driven by changes in point estimate. In other words, any post-mortem analysis of such significance reversals, using decompositions like (5), must necessarily find that the relative change in bias is larger than the relative change in standard errors. Overall, these results closely mirror empirical patterns of reversals of statistical insignificance observed in applied research, such as those documented by Lenz and Sahn (2021), and may offer a purely algebraic explanation for at least some of these patterns.

APPENDIX: DEFERRED PROOFS

PROOF OF THEOREM 2. From (6) and (7), the magnitude of the t-statistic for $H_0 : \lambda = \lambda_0$ can be written as a function of $R^2_{Y\sim Z|D\mathbf{X}}$ and $R^2_{D\sim Z|\mathbf{X}}$,

$$(9)\quad t(R^2_{Y\sim Z|D\mathbf{X}}, R^2_{D\sim Z|\mathbf{X}}) = \frac{|(\hat{\lambda}_r - \lambda_0) \pm \text{BF} \times \widehat{\text{se}}(\hat{\lambda}_r) \times \sqrt{\text{df}}|}{\widehat{\text{se}}(\hat{\lambda}_r) \times \text{SEF} \times \sqrt{\frac{\text{df}}{\text{df}-1}}}.$$

We wish to maximize $t(R^2_{Y\sim Z|D\mathbf{X}}, R^2_{D\sim Z|\mathbf{X}})$ under the posited bounds $R^2_{Y\sim Z|D\mathbf{X}} \leq R^{\max}_Y$ and $R^2_{D\sim Z|\mathbf{X}} \leq R^{\max}_D$.

That is, we want to solve the constrained maximization problem,

$$t^{\max}_{\boldsymbol{R^2}} = \max_{R^2_{Y\sim Z|D\mathbf{X}}, R^2_{D\sim Z|\mathbf{X}}} t(R^2_{Y\sim Z|D\mathbf{X}}, R^2_{D\sim Z|\mathbf{X}}) \tag{10}$$

such that $R^2_{Y\sim Z|D\mathbf{X}} \leq R^{\max}_Y$, $R^2_{D\sim Z|\mathbf{X}} \leq R^{\max}_D$. First notice that we should choose the direction of the bias that increases the magnitude of the difference $(\hat{\lambda} - \lambda_0)$. If $(\hat{\lambda}_r - \lambda_0) > 0$ then we should add the BF term in (9), whereas if $(\hat{\lambda}_r - \lambda_0) < 0$ then the bias should be subtracted. Both cases yield the same (simplified) objective function as argued below.

Suppose that $\hat{\lambda}_r > \lambda_0$. Then the absolute value from (10) can be dropped and the objective function becomes

$$\frac{\hat{\lambda}_r - \lambda_0 + \text{BF} \times \widehat{\text{se}}(\hat{\lambda}_r) \times \sqrt{\text{df}}}{\widehat{\text{se}}(\hat{\lambda}_r) \times \text{SEF} \times \sqrt{\frac{\text{df}}{\text{df}-1}}} = \frac{f_r \times \widehat{\text{se}}(\hat{\lambda}_r) + \text{BF} \times \widehat{\text{se}}(\hat{\lambda}_r)}{\widehat{\text{se}}(\hat{\lambda}_r) \times \text{SEF} \times \sqrt{\frac{1}{\text{df}-1}}}.$$

Now suppose that $\hat{\lambda}_r < \lambda_0$. We again have,

$$\frac{\lambda_0 - \hat{\lambda}_r + \text{BF} \times \widehat{\text{se}}(\hat{\lambda}_r) \times \sqrt{\text{df}}}{\widehat{\text{se}}(\hat{\lambda}_r) \times \text{SEF} \times \sqrt{\frac{\text{df}}{\text{df}-1}}} = \frac{f_r \times \widehat{\text{se}}(\hat{\lambda}_r) + \text{BF} \times \widehat{\text{se}}(\hat{\lambda}_r)}{\widehat{\text{se}}(\hat{\lambda}_r) \times \text{SEF} \times \sqrt{\frac{1}{\text{df}-1}}}.$$

Therefore, after some algebraic manipulation, the maximum t-value in (10) will be of the form,

$$t^{\max}_{\boldsymbol{R^2}} = \frac{f_r \sqrt{1 - R^{*2}_{D\sim Z|\mathbf{X}}} + \sqrt{R^{*2}_{Y\sim Z|D\mathbf{X}} R^{*2}_{D\sim Z|\mathbf{X}}}}{\sqrt{(1 - R^{*2}_{Y\sim Z|D\mathbf{X}})/(\text{df}-1)}}$$

where $R^{*2}_{Y\sim Z|D\mathbf{X}}$, $R^{*2}_{D\sim Z|\mathbf{X}}$ are the values of $R^2_{Y\sim Z|D\mathbf{X}}$ and $R^2_{D\sim Z|\mathbf{X}}$ that optimize (10).

We now find analytical expressions for the optimizers $R^{*2}_{Y\sim Z|D\mathbf{X}}$, $R^{*2}_{D\sim Z|\mathbf{X}}$. In what follows we write t_r for the objective function with the understanding that it is written in its modified form above. The partial derivative of $\frac{t_r}{\sqrt{\text{df}-1}}$ with respect to $R^2_{Y\sim Z|D\mathbf{X}}$ is

$$\frac{\partial t_r / \sqrt{\text{df}-1}}{\partial R^2_{Y\sim Z|D\mathbf{X}}} = \frac{f_r \sqrt{(1 - R^2_{D\sim Z|\mathbf{X}}) R^2_{Y\sim Z|D\mathbf{X}}} + \sqrt{R^2_{D\sim Z|\mathbf{X}}}}{2(1 - R^2_{Y\sim Z|D\mathbf{X}})^{\frac{3}{2}} \sqrt{R^2_{Y\sim Z|D\mathbf{X}}}}.$$

Since the partial of t_r with respect to $R^2_{Y\sim Z|D\mathbf{X}}$ is always positive, we have $R^{*2}_{Y\sim Z|D\mathbf{X}} = R^{\max}_Y$ unconditionally, i.e. $R^{*2}_{Y\sim Z|D\mathbf{X}}$ always lies on the boundary.

We now turn to the partial derivative of $\frac{t_r}{\sqrt{\text{df}-1}}$ with respect to $R^2_{D\sim Z|\mathbf{X}}$:

$$\frac{\partial t_r / \sqrt{\text{df}-1}}{\partial R^2_{D\sim Z|\mathbf{X}}} = \frac{-f_r \sqrt{R^2_{D\sim Z|\mathbf{X}}} + \sqrt{R^2_{Y\sim Z|D\mathbf{X}} (1 - R^2_{D\sim Z|\mathbf{X}})}}{2\sqrt{(1 - R^2_{Y\sim Z|D\mathbf{X}})(1 - R^2_{D\sim Z|\mathbf{X}}) R^2_{D\sim Z|\mathbf{X}}}}.$$

It is straight-forward to check that the second derivative of $\frac{t_r}{\sqrt{\text{df}-1}}$ is negative with respect to $R^2_{D\sim Z|\mathbf{X}}$. Thus, when attainable, the zero of the first partial derivative with respect to $R^2_{D\sim Z|\mathbf{X}}$ is a maximizer. Solving for the value that makes the first derivative zero yields:

$$R^{*2}_{D\sim Z|\mathbf{X}} = \frac{R^{\max}_Y}{f_r^2 + R^{\max}_Y}. \tag{11}$$

This interior point solution is only feasible when $f_r^2 \geq R^{\max}_Y (1 - R^{\max}_D)/R^{\max}_D$. Otherwise, if

$$f_r^2 < R^{\max}_Y (1 - R^{\max}_D)/R^{\max}_D, \tag{12}$$

then the partial with respect to $R^2_{D\sim Z|\mathbf{X}}$ is strictly positive for all $R^2_{D\sim Z|\mathbf{X}} \leq R^{\max}_D$ and so we obtain the boundary solution $R^{*2}_{D\sim Z|\mathbf{X}} = R^{\max}_D$. □

PROOF OF THEOREM 3. Let $\mathbf{Z}$ denote an $(n \times m)$ matrix of unobserved covariates and let $\hat{\boldsymbol{\gamma}}$ denote the coefficient vector of $\mathbf{Z}$. We are now interested in the long regression

$$Y = \hat{\lambda} D + \mathbf{X}\hat{\beta} + \mathbf{Z}\hat{\gamma} + \hat{\epsilon}. \tag{13}$$

Consider the $(n \times 1)$ vector $Z_L := \mathbf{Z}\hat{\gamma}$. The regression

$$Y = \hat{\lambda} D + \mathbf{X}\hat{\beta} + Z_L + \hat{\epsilon} \tag{14}$$

yields the same value for $\hat{\lambda}$; therefore, the bias induced by $\mathbf{Z}$ is equal to that induced by Z_L and $R^2_{Y\sim Z_L|D,\mathbf{X}} = R^2_{Y\sim \mathbf{Z}|D,\mathbf{X}}$. On the other hand, since $\hat{\gamma}$ is chosen solely to maximize $R^2_{Y\sim \mathbf{Z}|D,\mathbf{X}}$, we also have that $R^2_{D\sim Z_L|\mathbf{X}} \leq R^2_{D\sim \mathbf{Z}|\mathbf{X}}$. Now observe that the standard error formula from (7) holds for multivariate $\mathbf{Z}$ if we correctly adjust for the degrees of freedom. Further note that the bias of Z_L is a strictly increasing function of $R^2_{D\sim Z_L|X}$. Thus, the most adversarial choice of $\mathbf{Z}$ is such that $R^2_{D\sim Z_L|\mathbf{X}} = R^2_{D\sim \mathbf{Z}|\mathbf{X}}$. We can thus assess the maximum t-statistic of a matrix $\mathbf{Z}$ by considering that of a single adversarial vector Z_L and further adjusting for the degrees of freedom. □

PROOF OF COROLLARY 1. For any subset $\mathbf{X}_S$ of the columns of observed covariate matrix $\mathbf{X}$, recall that $R^2_{Y\sim \mathbf{X}_S|D} \leq R^2_{Y\sim \mathbf{X}|D}$ and $R^2_{D\sim \mathbf{X}_S} \leq R^2_{D\sim \mathbf{X}}$. Now apply the proof of Theorem 3 with the alteration that the constraint $R^2_{D\sim Z_L|\mathbf{X}} \leq R^2_{D\sim \mathbf{Z}|\mathbf{X}}$ is not necessarily tight, since Z_L may not be adversarial for a specific dataset. □

For all cases below, consider the following condition for significance:

$$t^*_{\alpha,\text{df}-1} \leq t^{\max}_{\boldsymbol{R^2}}. \tag{15}$$

PROOF OF THEOREM 4. First consider the case in which $f_r = 0$. This only happens if $\hat{\lambda}_r = \lambda_0$. Since here $R^2_{D\sim Z|\mathbf{X}} = 0$, the inclusion of Z does not alter the point estimate and we still obtain $\hat{\lambda} = \lambda_0$ after adjusting for Z. Therefore, the adjusted t-statistic will be zero regardless of the value of the standard error.

If $f_r > f^*_{\alpha,\text{df}-1}$ then we are already able to reject H_0 even if Z has zero explanatory power. Otherwise, given the constraints $R^2_{Y\sim Z|D\mathbf{X}} \leq \mathrm{XRVI}$ and $R^2_{D\sim Z|\mathbf{X}} = 0$, the expression for $t^{\max}_{\boldsymbol{R^2}}$ simplifies to

$$t^{\max}_{\mathrm{XRVI},0} = \max_{R^2_{Y\sim Z|D\mathbf{X}}} \frac{f_r}{\sqrt{1-R^2_{Y\sim Z|D\mathbf{X}}}}\sqrt{\mathrm{df}-1}$$

such that $R^2_{Y\sim Z|D\mathbf{X}} \leq \mathrm{XRVI}$. This is a strictly increasing function of $R^2_{Y\sim Z|D\mathbf{X}}$ and thus attains its maximum at $R^{*2}_{Y\sim Z|D\mathbf{X}} = \mathrm{XRVI}$. Thus solving for the minimum value of XRVI that satisfies (15) is equivalent to solving for XRVI at the equality. That is,

$$f^*_{\alpha,\mathrm{df}-1} = \frac{f_r}{\sqrt{1-\mathrm{XRVI}^0_\alpha}}.$$

Squaring and rearranging terms, we obtain

$$\mathrm{XRVI}^0_\alpha = 1 - \left(\frac{f_r}{f^*_{\alpha,\mathrm{df}-1}}\right)^2,$$

as desired. □

PROOF OF THEOREM 5. If $f_r > f^*_{\alpha,\mathrm{df}-1}$ then we are already able to reject H_0 even if Z has zero explanatory power, and thus the minimal strength to reject H_0 is zero. Otherwise, consider constraints $R^2_{Y\sim Z|D\mathbf{X}} \leq \mathrm{XRVI}$ and $R^2_{D\sim Z|\mathbf{X}} \leq 1$. From the proof of Theorem 2 we see that $t^{\max}_{\boldsymbol{R^2}}$ is an increasing function of XRVI. Thus solving for the minimum value of XRVI that satisfies (15) is equivalent to solving for XRVI at the equality. Notice that (12) is not satisfied here. We therefore plug in the interior point solution from Theorem 2 for $t^{\max}_{\boldsymbol{R^2}}$ and solve for XRVI^1_α. This results in the equation

$$f^*_{\alpha,\mathrm{df}-1} = \sqrt{\frac{f_r^2 + \mathrm{XRVI}^1_\alpha}{1-\mathrm{XRVI}^1_\alpha}}$$

which yields the solution

$$\mathrm{XRVI}^1_\alpha = \frac{f^{*2}_{\alpha,\mathrm{df}-1} - f_r^2}{1+f^{*2}_{\alpha,\mathrm{df}-1}},$$

as we wanted to show. □

THEOREM 7 (Closed-form expression for $\mathrm{XRVI}^{R^{\max}_D}_\alpha$). *The analytical expression for $XRVI^{R^{\max}_D}_\alpha$ is*

$$XRVI^{R^{\max}_D}_\alpha = \begin{cases} \frac{-b+s\sqrt{b^2-4ac}}{2a}, & \text{if } 0\leq f_r^2 < XRVI^1_\alpha\left(\frac{1-R^{\max}_D}{R^{\max}_D}\right) \\ XRVI^1_\alpha, & \text{if } XRVI^1_\alpha\left(\frac{1-R^{\max}_D}{R^{\max}_D}\right)\leq f_r^2 \leq f^{*2}_{\alpha,df-1}, \\ 0, & \text{if } f^{*2}_{\alpha,df-1} < f_r^2 \end{cases}$$

where

$$a = 1, \tag{16}$$

$$b = -2\left[\frac{f^{*2}_{\alpha,df-1}-(1-R^{\max}_D)f_r^2}{f^{*2}_{\alpha,df-1}+R^{\max}_D} + \frac{2f_r^2(1-R^{\max}_D)R^{\max}_D}{(f^{*2}_{\alpha,df-1}+R^{\max}_D)^2}\right], \tag{17}$$

$$c = \left[\frac{f^{*2}_{\alpha,df-1}-(1-R^{\max}_D)f_r^2}{f^{*2}_{\alpha,df-1}+R^{\max}_D}\right]^2, \tag{18}$$

$$XRVI^1_\alpha = \frac{f^{*2}_{\alpha,df-1}-f_r^2}{1+f^{*2}_{\alpha,df-1}}, \tag{19}$$

and $s\in\{-1,1\}$ is chosen to yield the valid quadratic root, i.e., $t^{\max}_{XRVI,R^{\max}_D} = t^*_{\alpha,df-1}$.

PROOF OF THEOREM 7. As argued in the proofs for Theorems 4 and 5, if $f^*_{\alpha,\mathrm{df}-1} < f_r$, then the minimum strength of Z necessary to attain significance is zero. Otherwise, solving for the minimum value of XRVI that satisfies (15) is equivalent to solving for XRVI at the equality.

Now let $f_r > 0$. Here we have two cases. Either both R^2 values reach the bound or the optimal value of $R^2_{D\sim Z|\mathbf{X}}$ is an interior point. For latter case, this means the constraint $R^2_{D\sim Z|\mathbf{X}}$ is not binding and thus $\mathrm{XRVI}^{R^{\max}_D}_\alpha$ should equal XRVI^1_α. Recall that t_r is concave down with respect to $R^{\max}_D$; therefore, the optimal value of $R^2_{D\sim Z|\mathbf{X}}$ is the interior point solution $R^{*2}_{D\sim Z|\mathbf{X}}$, as defined in (11), if and only if $R^{*2}_{D\sim Z|\mathbf{X}} < R^{\max}_D$. We can simplify this condition to be of the form,

$$\frac{\mathrm{XRVI}^1(1-R^{\max}_D)}{R^{\max}_D} < f_r^2.$$

It remains to solve for the case where both coordinates reach the bound. Here, the equality in (15) simplifies to

$$t^*_{\alpha,\mathrm{df}-1} = \frac{f_r\sqrt{1-R^{\max}_D}+\sqrt{R^{\max}_D \mathrm{XRVI}}}{\sqrt{(1-\mathrm{XRVI})/(\mathrm{df}-1)}}. \tag{20}$$

We now solve for XRVI in (20) by squaring both sides and taking the valid root of the quadratic equation. The simplified quadratic form is

$$\mathrm{XRVI}^2 - \left[\frac{2(f^{*2}_{\alpha,\mathrm{df}-1}-(1-R^{\max}_D)f_r^2)}{f^{*2}_{\alpha,\mathrm{df}-1}+R^{\max}_D} - \frac{4f_r^2(1-R^{\max}_D)R^{\max}_D}{(f^{*2}_{\alpha,\mathrm{df}-1}+R^{\max}_D)^2}\right]\mathrm{XRVI} \tag{21}$$

$$+\left[\frac{f^{*2}_{\alpha,\mathrm{df}-1}-(1-R^{\max}_D)f_r^2}{f^{*2}_{\alpha,\mathrm{df}-1}+R^{\max}_D}\right]^2 = 0. \tag{22}$$

The expressions for a,b,c in (16)-(18) immediately follow from Sridharacharya-Bhaskara's formula for quadratic equations. Finally, we note that if $f_r = 0$ and $R^{\max}_D > 0$ then $\mathrm{XRVI}^{R^{\max}_D}$ solution is valid. If $f_r = 0$ and $R^{\max}_D = 0$ then we are in the XRVI^0 case such that, by Theorem 4, we cannot overturn an insignificant result. □

PROOF OF THEOREM 6. The derivation of RVI_α follows that of $\mathrm{XRVI}^{R^{\max}_D}_\alpha$ very closely (see proof of Theorem 7). The only difference is that when $f_r^2 < 1-\mathrm{RVI}$, the equality in (15) simplifies to

$$t^*_{\alpha,\mathrm{df}-1} = \frac{f_r\sqrt{1-\mathrm{RVI}}+\mathrm{RVI}}{\sqrt{1-\mathrm{RVI}}} \times \sqrt{\mathrm{df}-1}$$

which is equivalent to

$$f^*_{\alpha,\mathrm{df}-1} = f_r + \frac{\mathrm{RVI}}{\sqrt{1-\mathrm{RVI}}}. \tag{23}$$

Let $f_\Delta = f^*_{\alpha,\mathrm{df}-1} - f_r$. Then (23) further reduces to a quadratic function of RVI with positive root

$$\mathrm{RVI} = \frac{1}{2}\left(\sqrt{f_\Delta^4 + 4f_\Delta^2} - f_\Delta^2\right).$$

It remains to show that (12), i.e. $f_r^2 \geq 1 - \mathrm{RVI}$, is equivalent to

$$f_r > \frac{1}{f^*_{\alpha,\mathrm{df}-1}}. \tag{24}$$

This is equivalent to showing that RVI_α is given by the interior point solution if and only if (24) holds. To see why, recall that for the interior point solution,

$$R_Y^{\max} = R_D^{\max} = \mathrm{RVI}_\alpha = \frac{f^{*2}_{\alpha,\mathrm{df}-1} - f_r^2}{1 + f^{*2}_{\alpha,\mathrm{df}-1}}. \tag{25}$$

Plugging (25) into (12), we have

$$f_r^2 \geq 1 - \mathrm{RVI}_\alpha = 1 - \frac{f^{*2}_{\alpha,\mathrm{df}-1} - f_r^2}{1 + f^{*2}_{\alpha,\mathrm{df}-1}}$$

which reduces to $f^*_{\alpha,\mathrm{df}-1} \geq \frac{1}{f_r}$. □

ASYMPTOTIC DISTRIBUTION OF $R^2_{D\sim Z|\mathbf{X}}$. First consider the case without observed covariates $\mathbf{X}$. Under i.i.d sampling, the asymptotic distribution of the sample correlation $R_{D\sim Z}$ is derived in Ferguson (2017, Theorem 8, p. 52) to be

$$\sqrt{df}\,(R_{D\sim Z} - \rho) \xrightarrow{d} \mathcal{N}(0, \gamma^2)$$

where here $\xrightarrow{d}$ denotes convergence in distribution, ρ is the *population* correlation coefficient of D and Z, and

$$\gamma^2 = c_1\rho^2 - c_2\rho + \frac{\mathrm{Var}((D - \mathbb{E}[D])(Z - \mathbb{E}[Z]))}{\mathrm{Var}(D)\,\mathrm{Var}(Z)},$$

where constants c_1 and c_2 depend on higher order moments of D and Z, and $\mathrm{Var}(\cdot)$ denotes *population* variance. If D is randomized, we have that D is independent of Z by design. Thus $\rho = 0$ and $\mathrm{Var}((D - \mathbb{E}[D])(Z - \mathbb{E}[Z])) = \mathrm{Var}(D)\,\mathrm{Var}(Z)$. This simplifies the expression of the asymptotic variance γ^2 to 1, and we have

$$\sqrt{df}\,R_{D\sim Z} \xrightarrow{d} \mathcal{N}(0,1), \quad \text{and} \quad df\,R^2_{D\sim Z} \xrightarrow{d} \chi^2_1.$$

To extend the argument to the sample partial correlation $R_{D\sim Z|\mathbf{X}}$, first note that it can be rewritten using the FWL Theorem (Frisch and Waugh, 1933; Lovell, 1963) as the sample correlation $R_{D\sim Z|\mathbf{X}} = \mathrm{cor}(\tilde{Z}, \tilde{D})$, where $\tilde{Z}$ and $\tilde{D}$ are the sample residuals of the regression of Z and D on $\mathbf{X}$, i.e, $\tilde{Z} := Z - \mathbf{X}\hat{\theta}$ and $\tilde{D} := D - \mathbf{X}\hat{\delta}$, and $\hat{\theta}$ and $\hat{\delta}$ are the respective OLS coefficient estimates, which are asymptotically normal. Now define the *population* counterparts, $\check{Z} := Z - \mathbf{X}\theta$, $\check{D} := D - \mathbf{X}\delta$, where we replace sample estimates with their corresponding population values. We note that estimation errors on $\hat{\theta}$ and $\hat{\delta}$ do not affect the asymptotic distribution of $\mathrm{cor}(\tilde{Z}, \tilde{D})$, which is the same as that of $\mathrm{cor}(\check{Z}, \check{D})$—this can be verified by applying standard results in large sample theory to the covariances (and variances) of sample residuals, such as Boos and Stefanski (2013, Theorem 5.28, p. 249). Now $\mathrm{cor}(\check{Z}, \check{D})$ is again a simple bivariate correlation, and we can directly apply the result of Ferguson (2017) above. Note this result does not rely on any parametric distributional assumptions on $\check{D}$ and $\check{Z}$, except for requiring that the relevant moments are finite. □

FUNDING

This work is supported in part by the Royalty Research Fund at the University of Washington, and by the National Science Foundation under Grant No. 2417955.

REFERENCES

AMLANI, S. and COLLITT, S. (2022). The Impact of Vote-By-Mail Policy on Turnout and Vote Share in the 2020 Election. *Election Law Journal: Rules, Politics, and Policy* **21** 135–149.

BOOS, D. D. and STEFANSKI, L. A. (2013). *Essential statistical inference: theory and methods* **591**. Springer.

BRODEUR, A., LÉ, M., SANGNIER, M. and ZYLBERBERG, Y. (2016). Star wars: The empirics strike back. *American Economic Journal: Applied Economics* **8** 1–32.

CINELLI, C., FORNEY, A. and PEARL, J. (2022). A crash course in good and bad controls. *Sociological Methods & Research* 00491241221099552.

CINELLI, C. and HAZLETT, C. (2020). Making Sense of Sensitivity: Extending Omitted Variable Bias. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **82** 39–67.

CINELLI, C. and HAZLETT, C. (2022). An omitted variable bias framework for sensitivity analysis of instrumental variables. *Available at SSRN 4217915*.

FERGUSON, T. S. (2017). *A course in large sample theory*. Routledge.

FRISCH, R. and WAUGH, F. V. (1933). Partial time regressions as compared with individual trends. *Econometrica: Journal of the Econometric Society* 387–401.

LENZ, G. S. and SAHN, A. (2021). Achieving statistical significance with control variables and without transparency. *Political Analysis* **29** 356–369.

LOVELL, M. C. (1963). Seasonal adjustment of economic time series and multiple regression analysis. *Journal of the American Statistical Association* **58** 993–1010.

SIMONSOHN, U., NELSON, L. D. and SIMMONS, J. P. (2014). P-curve: a key to the file-drawer. *Journal of experimental psychology: General* **143** 534.

VIVALT, E. (2019). Specification searching and significance inflation across time, methods and disciplines. *Oxford Bulletin of Economics and Statistics* **81** 797–816.