How Much Should We Trust Instrumental Variable Estimates in Political Science? Practical Advice Based on 67 Replicated Studies*

Apoorva Lal (Stanford) Mac Lockhart (UCSD)
Yiqing Xu (Stanford) Ziwen Zu (UCSD)

First Version: July 10, 2021
This Version: April 19, 2023

Abstract

Instrumental variable (IV) strategies are widely used in political science to establish causal relationships, but the identifying assumptions required by an IV design are demanding, and assessing their validity remains challenging. In this paper, we replicate 67 papers published in three top political science journals from 2010-2022 and identify several concerning patterns. First, researchers often overestimate the strength of their instruments due to non-i.i.d. error structures such as clustering. Second, the commonly used t-test for two-stage-least-squares (2SLS) estimates frequently underestimates uncertainty. Using more robust inferential methods, we find that about 19-30% of the 2SLS estimates in our sample are underpowered. Third, in most replicated studies, 2SLS estimates are significantly larger than ordinary-least-squares estimates, with their ratio negatively correlated with instrument strength in studies with non-experimentally generated instruments, suggesting potential violations of unconfoundedness or exclusion restriction. We provide a checklist and software to help researchers avoid these pitfalls and improve their practice.

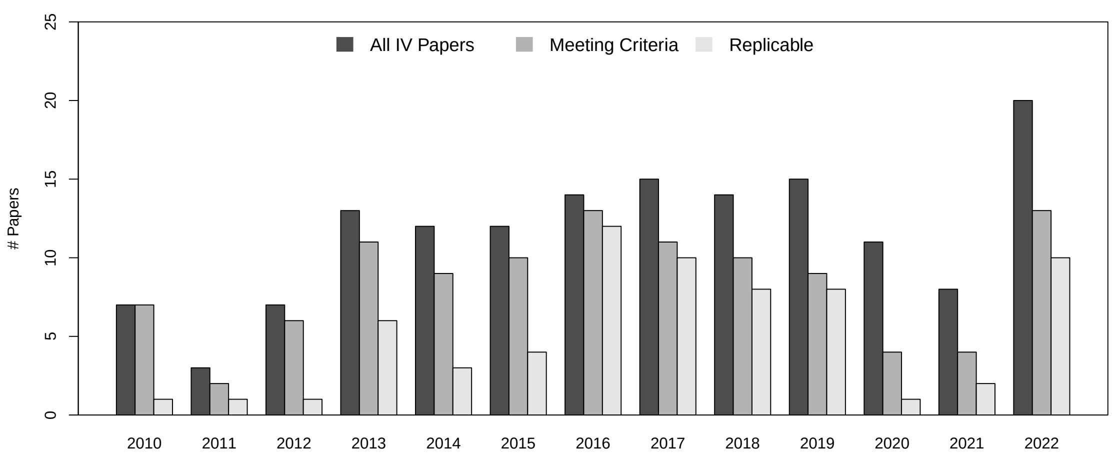
Keywords: instrumental variables, two-stage-least-squared, replications, weak instrument, exclusion restriction,

*Apoorva Lal, PhD Candidate, Stanford University; Email: apoorval@stanford.edu. Mac Lockhart, PhD Candidate, University of California, San Diego; Email: mwlockha@ucsd.edu. Yiqing Xu (corresponding author), Assistant Professor, Stanford University; Email: yiqingxu@stanford.edu Ziwen Zu, PhD Student, University of California, San Diego; Email: zzu@ucsd.edu. We thank Te Bao, Daniel Chen, Gary Cox, Charles Crabtree, Ted Enamorado, Hanming Fang, Avi Feller, Don Green, Justin Grimmer, Anna Grzymala-Busse, Jens Hainmueller, David Laitin, Justin McCrary, Jacob Montgomery, Doug Rivers, Henrik Sigstad, Brandon Stewart, Arthur Yu, Xiang Zhou, and seminar participants at Stanford University, Washington University in St. Louis, APSA 2021, and Polmeth 2021 for extremely valuable comments.

1. Introduction

The instrumental variable (IV) approach is a commonly used empirical method in the social sciences, including political science, for establishing causal relationships. It is often used when selection on observables is implausible, experimentation is infeasible or unethical, and rule-based assignments that allow for sharp regression-discontinuity (RD) designs are not available. In recent years, there has been a growing number of papers published in top political science journals, such as the *American Political Science Review* (APSR), *American Journal of Political Science* (AJPS), and *Journal of Politics* (JOP), that use IV as a primary causal identification strategy. This trend can be traced back to the publication of the textbook *Mostly Harmless Econometrics* (Angrist and Pischke, 2008), which popularized the modern interpretation of IV designs, and Sovey and Green (2011), which clarifies the assumptions required by an IV approach and provides a useful checklist for political scientists.

FIGURE 1. IV PAPERS PUBLISHED IN THE *APSR*, *AJPS*, AND *JOP*



Note: Our criteria rule out IV models appearing in the online appendix only, in dynamic panel settings, with multiple endogenous variables, and with nonlinear link functions. Non-replicability is primarily due to a lack of data and/or coding errors.

Despite its popularity, some researchers have questioned the validity of the IV approach, noting that two-stage least-squares (2SLS) estimates are often much larger in magnitude than "naïve" ordinary-least-squares (OLS) estimates, even when the main concern with the latter is upward omitted-variables bias.[1] Others have raised concerns about the validity of inferential methods used for 2SLS estimation (e.g. Lee et al., 2022; Young, 2022).

These observations motivate our systematic examination of the use of IVs in the empirical political science literature. We set out to replicate all papers published in the APSR, AJPS, and JOP during the past thirteen years (2010-2022) that use an IV design with a single endogenous variable as one of the main identification strategies.[2] Out of 114 papers meeting this criterion, 71 have complete replication materials online, which is itself a troubling pattern. We successfully replicate at least one of the main IV results for 67 of the 71 remaining papers, with three papers having two separate IV designs producing separate 2SLS results.

Using data from these 70 IV designs, we conduct a programmatic replication exercise and find three troubling patterns. First, a significant number of IV designs in political science either do not report the first-stage partial F statistic or overestimate it by failing to adjust standard errors (SEs) for factors such as heteroskedasticity, serial correlation, or clustering structure. Using the effective F-statistic (Olea and Pflueger, 2013), we find that at least 11% of the published IV studies rely on what econometricians call "weak instruments," the consequences of which have been well-documented in the literature (see Andrews, Stock and Sun (2019) for a comprehensive review).

A second related issue is statistical inference. We find that a considerable number of IV designs are underpowered, with almost all studies relying on t-tests based on analytic standard errors and traditional critical values (such as 1.96 for statistical significance at the

[1]For example, in the 2016 National Bureau of Economic Research–Political Economy Meeting, following a presentation of a study using an IV approach, the late political economist Alberto Alesina asked the audience: "How come 2SLS estimates are always five times bigger than OLS estimates in political economy?"

[2]Focusing on design with a single endogenous variable allows us to calculate the correlation coefficient between the treatment and the predicted treatment and apply powerful tools such as the Anderson-Rubin (AR) test and the tF test (when there is only a single instrument). Moreover, we find it difficult to justify the exclusion restriction in a multiple-treatment-multiple-instrument setting in the first place.

5% level) to make inferences about the 2SLS coefficients. When we use bootstrapping procedures, the AR test, or the tF procedure, an F-statistic-dependent t-test (Lee et al., 2022), to perform hypothesis testing, we find that, depending on the method employed, 19-30% of designs cannot reject the null hypothesis of no effect at the 5% level, whereas the number based on the SEs or p-values reported in the original papers is only 10%. This suggests that inferences based on traditional t-tests may not accurately reflect the uncertainties in 2SLS estimates in a significant portion of cases.

Finally, our replications corroborate evidence from economics and finance that the 2SLS estimates are often much bigger in magnitude than the OLS estimates obtained from regressing the outcome on the potentially endogenous treatment variables and covariates (Jiang, 2017). In 68 out of the 70 designs (97%), the 2SLS estimates are bigger than the OLS estimates in magnitude; among them, 24 (34%) are at least five times bigger. This is alarming because, in an IV design with observational data, researchers often say that they are most concerned about the upward bias of the treatment effect estimates produced by naïve OLS. Even after we exclude 15 papers that explicitly claim to expect downward biases in OLS estimates, the percentages remain high (96% and 35%, respectively).

The first two patterns may be due to researchers' unfamiliarity with recent development in the IV literature, such as the effective F statistic and the tF test, or under-utilization of inferential procedures robust to weak instruments, such as the AR test. Therefore, researchers can avoid these problems by adopting better practices. The third finding, however, is the most concerning. We cannot explain it with weak instruments alone because at least in the case of i.i.d. errors, when instruments are exogenous, weak instruments bias 2SLS estimates toward OLS estimates in finite samples (Bound, Jaeger and Baker, 1995). But what we observe is the opposite: The ratio between the magnitudes of the 2SLS and OLS estimates is strongly negatively correlated with the strength of the first stage among studies that use non-experimental instruments, and the relationship is almost nonexistent among studies with experimental instruments. We suspect that this is primarily driven by a combination of weak instruments and failure of exogeneity, although other mechanisms such as publication bias, heterogeneous treatment effects (HTE), and measurement error may also contribute.

What do these findings mean for empirical IV studies in political science? First, traditional t tests for the 2SLS estimates (especially those based on classic analytic SEs) mask the fact that most IV results are highly uncertain, which likely leads to selective reporting and publication bias. Second, and more importantly, many of the 2SLS estimates likely suffer from large biases due to failures of unconfoundedness or exclusion restriction and hence are not credible. Although we cannot definitively say which estimates are problematic, the underlying issue seems to prevail in the IV literature. However, the goal of this paper is not to discredit existing IV studies or dissuade researchers from ever using the IV method. On the contrary, we want to caution researchers against ad-hoc justifications for IVs in observational studies and provide practical advice to improve future practices. This includes accurately quantifying instrument strength and 2SLS estimate uncertainties, as well as conducting additional analysis, such as placebo tests, to corroborate the identifying assumptions.

Our work builds on a growing literature evaluating IV strategies in social sciences and offering methods to improve empirical practice. Notable studies include Young (2022), which finds IV estimates to be more sensitive to outliers and conventional t-tests to understate uncertainties; Jiang (2017), which observes larger IV estimates in finance journals and attributed this to exclusion restriction violations and weak instruments; Mellon (2020), which emphasizes the vulnerability of weather instruments; Dieterle and Snell (2016), which develops a quadratic over-identification test and discovered significant non-linearities in the first stage regression; Felton and Stewart (2022), which finds unstated assumptions and a lack of weak-instrument robust tests in top sociology journals; and Cinelli and Hazlett (2022), which proposes a sensitivity analysis for IV designs in an omitted variable bias framework.

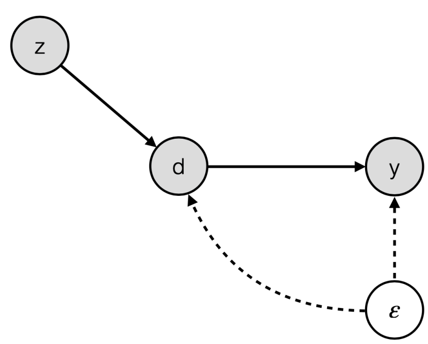
Our study is the first large-scale replication effort focusing on IV designs and the first to link the discrepancy between 2SLS and OLS estimates to weak instruments using extensive replication data across social sciences.

2. Theoretical Refresher

In this section, we offer a brief overview of the IV approach, including the setup, the identifying assumptions, as well as the 2SLS estimators. We then discuss potential pitfalls and survey several inferential methods. To cover the vast majority of IV studies in political science, we adopt a traditional constant treatment effect approach to IV designs, which imposes a set of parametric assumptions. For example, 51 (73%) designs in our replicated sample employ continuous treatment variables and make no reference to treatment effect heterogeneity, hence, they are ill-suited for the local average treatment effect (LATE) approach (Angrist, Imbens and Rubin, 1996).

For simplicity, we do not include additional exogenous controls in the discussion without loss of generality. This is because, by the Frisch-Waugh-Lovell theorem, we can remove them by regressing the outcome, treatment, and IVs on the controls and using the residuals for all subsequent analyses.

Figure 2. A Directed Acyclic Graph of an IV Design



Apart from the "canonical" use of IVs in addressing non-compliance in experimental encouragement designs, we observe that in the majority of the papers we review, researchers use IVs in observational settings to establish causality between a single treatment variable d and an outcome variable y. The basic idea of this approach is to use an instrument z to isolate "exogenous" variation in d (i.e., the variation in d that is not related to potential confounders) and estimate its causal effect on y. Figure 1 illustrates the directed acyclic graph for an IV design, where ε denotes the error term that captures all unexplained variations in y. The figure depicts that because d and ε are correlated, an observed correlation between d and y does not identify the causal effect of d on y. It also shows that an IV approach relies on three crucial assumptions: (1) the *relevance* of the instrument, which is directly testable, meaning that z is correlated with d; (2) the *unconfoundedness* assumption, which states that z is quasi-randomly assigned, and (3) the *exclusion restriction*, which posits that z does not have a direct effect on y beyond the channel through d.

2.1. Estimation Strategies

Imposing a set of parametric assumptions, we define a system of simultaneous equations:

$$\text{Structural equation:} \quad y = \alpha + \tau d + \varepsilon \tag{2.1}$$

$$\text{First-stage equation:} \quad d = \pi_0 + \pi_1 z + \nu \tag{2.2}$$

in which y is the outcome variable, d is a scalar treatment variable; z is a vector of instruments for d; τ captures the (constant) treatment effect and is the key quantity of interest. Equations (2.1) and (2.2) are referred to as the structural equation and the first-stage equation, respectively. The error terms ε and ν in the two equations may be correlated.

The endogeneity problem for τ in Equation (2.1) arises when d and ε are correlated, which renders $\hat{\tau}_{OLS}$ from a naïve OLS regression of y on d inconsistent. The endogeneity problem may be due to one of the following reasons: (1) unmeasured omitted variables that are correlated with both y and d; (2) measurement error in d, or (3) simultaneity or reverse causality, which means y may also affect d. The IV approach addresses this problem by taking advantage of the exogenous variation in d brought by z. Substituting d in Equation (2.1) using Equation (2.2), we have the reduced form equation:

$$\text{Reduced form:} \quad y = \underbrace{(\alpha + \tau\pi_0)}_{\gamma_0} + \underbrace{(\tau\pi_1)}_{\gamma_1} z + (\tau\nu + \varepsilon). \tag{2.3}$$

Substitution establishes that $\gamma_1 = \tau\pi_1$, rearranging yields $\tau = \frac{\gamma_1}{\pi_1}$ (assuming that we only use one instrument, but the intuition carries over to cases with multiple instruments). The IV estimate, therefore, is the ratio of the reduced-form and first-stage coefficients. To identify τ, we make the following assumptions (Greene, 2003, Chapter 12).

Assumption 1 (Relevance).

$\pi_1 \neq 0$. This assumption requires that the IVs can predict the treatment variable, and is therefore equivalently stated as $d \not\perp\!\!\!\perp z$.

Assumption 2 (Exogeneity).

$\text{Cov}(z, \varepsilon) = 0$ and $\mathbb{E}[\varepsilon] = 0$. Assumption 2 is satisfied when unconfoundedness and the exclusion restriction are satisfied. However, without additional structural assumptions. failures of unconfoundedness and the exclusion restriction are observationally equivalent, therefore, we do not distinguish them in the analysis and diagnostics.

Under Assumptions 1 and 2, the 2SLS estimator is shown to be consistent for the structural parameter τ. Consider a sample of N observations. We can write $\mathbf{d} = (d_1, d_2, \cdots, d_N)'$ and $\mathbf{y} = (y_1, y_2, \cdots, y_N)'$ as $(N \times 1)$ vectors of the treatment and outcome data, and $\mathbf{z} = (z_1, z_2, \cdots, z_N)'$ as $(N \times p_z)$ matrix of instruments in which p_z is the number of instruments. To simplify mathematics, we residualize original $\mathbf{d}$, $\mathbf{y}$, and each column of $\mathbf{z}$ against the exogenous covariates, obtaining $\mathbf{y}$, $\mathbf{d}$, and $\mathbf{z}$, respectively. The 2SLS estimator is written as follows:

$$\hat{\tau}_{\text{2SLS}} = (\mathbf{d}'\mathbf{P}_z\mathbf{d})^{-1}\,\mathbf{d}'\mathbf{P}_z\boldsymbol{y} \tag{2.4}$$

in which $\mathbf{P}_z = \mathbf{z}\left(\mathbf{z}'\mathbf{z}\right)^{-1}\mathbf{z}'$ is the hat-maker matrix from the first stage which projects the endogenous treatment variable $\mathbf{d}$ into the column space of $\mathbf{z}$, thereby preserving only the exogenous variation in $\mathbf{d}$ that is uncorrelated with ε. This formula permits the use of more than one instrument, in which case the model is said to be "overidentified." The 2SLS estimator belongs to a class of generalized method of moments (GMM) estimators taking advantage of the moment condition $\mathbb{E}[z\varepsilon] = 0$, including the two-step GMM (Hansen, 1982) and limited information maximum likelihood (LIML) estimators (Anderson, Kunitomo and Sawa, 1982). We use the 2SLS estimator throughout the replication exercise because of its simplicity and because every single paper in our replication sample uses it in at least one specification.

When the model is exactly identified, i.e., the number of treatment variables equals the number of instruments, the 2SLS estimator can be simplified as the IV estimator: $\hat{\tau}_{\text{2SLS}} = \hat{\tau}_{\text{IV}} = \left(\mathbf{z}'\mathbf{d}\right)^{-1}\mathbf{z}'\mathbf{y}$. In the case of one instrument and one treatment, the 2SLS estimator can also be written as a ratio of two sample covariances: $\hat{\tau}_{2SLS} = \hat{\tau}_{IV} = \frac{\hat{\gamma}_1}{\hat{\pi}_1} = \frac{\widehat{\text{Cov}}(\mathbf{y},\mathbf{z})}{\widehat{\text{Cov}}(\mathbf{d},\mathbf{z})}$, which illustrates that the 2SLS estimator is a ratio between reduced-form and first-stage coefficients in this special case. This further simplifies to a ratio of the difference in means when z is binary, which is called a Wald estimator.

2.2. Potential Pitfalls in Implementing an IV Strategy

The challenges with 2SLS estimation and inference are often due to the violation of the two identifying assumptions. These difficulties can result in (1) significant uncertainty around 2SLS estimates and size distortion for t tests due to weak instruments even when Assumption 2 is valid; and (2) potentially larger biases in 2SLS estimates compared to OLS estimates when both assumptions are violated.

Inferential problem due to weak instruments. Since the IV coefficient is a ratio, the weak instrument problem is a "divide-by-zero" problem, which arises when $\text{Cov}(z, x) \approx 0$ (i.e., when Assumption 1 is violated). The instability of ratio estimators like $\widehat{\tau}_{\text{2SLS}}$ when the denominator is approximately zero has been extensively studied going back to Fieller (1954). The conventional wisdom in the past two decades has been that the first-stage partial F statistic needs to be bigger than 10, and it should be clearly reported (Staiger and Stock, 1997). As a rule of thumb, the original cutoff is chosen based on simulation results to meet two criteria under i.i.d. errors: (1) in the worst case, the bias of the 2SLS estimator does not exceed 10% of the bias of the OLS estimator, and (2) a t-test based on the 2SLS estimator with a size of 5% does not lead to size over 15%.

The literature has discussed at least three issues caused by weak instruments when Assumption 2 is valid. First, under i.i.d. errors, a weak first stage exacerbates the finite-sample bias of the 2SLS estimator toward the inconsistent OLS estimator, thereby reproducing the endogeneity problem that an IV design was meant to solve (Staiger and Stock, 1997)[3]. Second, the 2SLS estimates become very imprecise.[4] A third and related issue is that the tests are of the wrong size and the t-statistics don't follow a t-distribution (Nelson and Starz, 1990). Issues relating to imprecision and test-statistic size arise from the fact that the distribution of $\hat{\tau}$ is derived from its linear approximation of $\hat{\tau}$ in $(\hat{\gamma}, \hat{\pi})$, wherein normality of the two OLS coefficients implies the normality of their ratio. However, this normal approximation breaks down when $\hat{\pi} \approx 0$. Moreover, this approximation failure cannot generally be rectified by bootstrapping (Andrews and Guggenberger, 2009), although Young (2022) argues that it nevertheless allows for improved inference when outliers are present. Overall,

[3]The 2SLS estimator may not have a mean when the first stage is weak, its median is centered around the OLS coefficient (Hirano and Porter, 2015)

[4]To illustrate, a commonly used variance estimator for $\hat{\tau}_{IV}$ can be written as: $\hat{\mathbb{V}}(\hat{\tau}_{IV}) \approx \frac{\hat{\Sigma}^2}{\sum_{i=1}^{N}(x_i - \overline{x})^2} \frac{1}{R_{xz}^2} = \hat{\mathbb{V}}(\hat{\tau}_{OLS}) \frac{1}{R_{xz}^2}$ in which $\hat{\sigma}^2$ is a variance estimator for the error term and R_{xz}^2 is the R-squared from the first stage. The estimated variance is mechanically larger than the estimated variance of the OLS estimator as long as $R_{xz}^2 < 1$. It is decreasing in R_{xz}^2, i.e. stronger instruments produce more precise IV estimates.

valid IV inference relies crucially on correctly identifying strong IVs.

In general, there are two approaches to conducting inference in an IV design: pretesting and direct testing. The pretesting approach involves using an F statistic to test the first stage strength, and if it exceeds a certain threshold (e.g., $F > 10$), proceeding to test the null hypothesis about the treatment effect (e.g., $\tau = 0$). In contrast, the direct testing approach does not rely on passing a pretest. Nearly all reviewed studies employ the pretesting approach. We examine four methods for statistical inference in IV designs, with the first three related to pretesting and the last one being a direct test.

First, Olea and Pflueger (2013) propose the effective F statistic for both just-identified and over-identified settings and accommodates robust or cluster-robust SEs. The effective F is a scaled version of the first-stage F statistic and is computed as $F_{\text{Eff}} = \hat{\pi}'\hat{Q}_{\text{ZZ}}\hat{\pi}/\text{tr}(\hat{\Sigma}_{\pi\pi}\hat{Q}_{\text{ZZ}})$, where $\hat{\Sigma}_{\pi\pi}$ is the variance-covariance matrix of the first stage regression, and $\hat{Q}_{\text{ZZ}} = \frac{1}{N}\sum_{i=1}^{N} z_i z_i'$. In just-identified cases, F_{Eff} is the same as F statistics based on robust or cluster-robust SEs. The authors derive the critical values for F_{Eff} and note that the statistic and corresponding critical values are identical to the better-known robust F statistic $\hat{\pi}\hat{\Sigma}_{\pi\pi}^{-1}\hat{\pi}$ and corresponding Stock and Yogo (2005) critical values. $F_{\text{Eff}} > 10$ is shown to be a reasonable rule of thumb under heteroskedasticity in simulations (Olea and Pflueger, 2013; Andrews, Stock and Sun, 2019).

Second, Young (2022) recommends researchers report two types of bootstrap confidence intervals (CIs), *bootstrap-c* and *bootstrap-t*, for $\hat{\tau}_{2SLS}$ under non-i.i.d. errors with outliers, which is common in social science settings. This involves B replications of the following procedure: (1) sample n triplets $(y_i^*, d_i^*, \mathbf{z}_i^*)$ independently and with replacement from the original sample (with appropriate modifications for clustered dependence) and (2) compute the $\hat{\tau}_{\text{2SLS}}$ coefficient and SE, as well as the corresponding test statistic $t^* = \hat{\tau}^*_{\text{2SLS}}/\hat{\text{SE}}(\hat{\tau}^*_{\text{2SLS}})$ on each replication. The *bootstrap-c* method calculates the CIs by taking the $\alpha/2$ and $(1-\alpha/2)$ percentiles of the bootstrapped 2SLS coefficient $\hat{\tau}^*_{\text{2SLS}}$, while the *bootstrap-t* method calculates the percentile-t refined CIs by plugging in the $\alpha/2$ and $(1-\alpha/2)$ percentile of bootstrapped t statistics $t^*_{\alpha/2}$ and $t^*_{1-\alpha/2}$ into the expression $\hat{\tau}_{\text{2SLS}} \pm t^*_{\alpha|1-\alpha}\hat{\text{SE}}(\hat{\tau}^*_{\text{2SLS}})$. Hall and Horowitz (1996) show that *bootstrap-t* achieves an asymptotic refinement over *bootstrap-c*.[5]

Third, in just-identified single treatment settings, Lee et al. (2022) propose the tF procedure that smoothly adjusts the $t-$ratio inference based on the first-stage F statistic, which improves upon the ad-hoc screening rule of $F > 10$. The adjustment factor applied to 2SLS SEs is based on the first stage $t-$ratio $\hat{f} := \hat{\pi}/\sqrt{\hat{\mathbb{V}}(\hat{\pi})}$, with the first stage $\hat{F} = \hat{f}^2$, and relies on the fact that the distortion from employing the standard 2SLS t-ratio $\hat{t} := \hat{\tau}/\sqrt{\hat{\mathbb{V}}(\hat{\tau})}$ can be quantified in terms of $\hat{F}$ and $AR-$statistic, which gives rise to a set of critical values for a given pair of $\hat{t}$ and $\hat{F}$. The authors also show that, if no adjustment is made to the t-test's critical value (e.g., using 1.96 as the threshold for 5% statistical significance), a first stage $\hat{F}$ of 104.7 is required to guarantee a correct size of 5% for a two-sided t-test for the 2SLS coefficient.

Finally, where there is one endogenous treatment variable, the AR procedure, which is essentially an F test on the reduced form, is a direct inferential method robust to weak instruments (Anderson and Rubin, 1949; Chernozhukov and Hansen, 2008). Without loss of generality, assume that we are interested in testing the null hypothesis that $\tau = 0$, which then implies that the reduced form coefficient from regressing y on $\mathbf{z}$, $\gamma_1 = 0$. This motivates the following procedure: given a set $\mathcal{T}$ of potential values for $\widetilde{\tau}$, for each value $\widetilde{\tau}$, construct $\widetilde{y} = y - d\widetilde{\tau}$, and regress $\widetilde{y}$ on $\mathbf{z}$ to obtain a point estimate $\widetilde{\gamma}$ and (robust, or cluster robust) covariance matrix $\widetilde{\mathbb{V}}(\widetilde{\gamma})$, and construct a Wald statistic $\widetilde{W}_s(\widetilde{\gamma}) := \widetilde{\gamma}'\widetilde{\mathbb{V}}(\widetilde{\gamma})^{-1}\widetilde{\gamma}$. Then, the AR CI is the set of $\widetilde{\gamma}$ such that $\widetilde{W}_s(\widetilde{\gamma}) \leq c(1-p)$ where $c(1-p)$ is the $(1-p)^{\text{th}}$ percentile of the χ^2_1 distribution. The AR test not only requires no pretesting but is also shown to be the uniformly most powerful unbiased test in the just-identified case (Moreira, 2009). However,

[5]We use the percentile method instead of bootstrapped SEs because the t-test based on the latter may be overly conservative (Hahn and Liao, 2021).

it is not as commonly used as procedures that involve pretesting, possibly because researchers are more accustomed to using t-tests than F/Wald tests and reporting SEs rather than CIs.

Bias amplification and the failure of Assumption 2. When the number of instruments is bigger than the number of endogenous treatments, researchers can use an overidentification test to gauge the plausibility of Assumption 2 (Arellano, 2002). However, such a test is often underpowered and has bad finite sample properties (Davidson and MacKinnon, 2015). In just-identified cases, Assumption 2 is not directly testable. When combined with weak instruments, even small violations of Assumption 2 can produce inconsistency. This is because plim $\hat{\tau}_{IV} = \tau + \frac{\text{Cov}(z,\varepsilon)}{\text{Cov}(z,d)}$. When $\text{Cov}(z,d) \approx 0$, even small violations of exogeneity, i.e., $\text{Cov}(z,\varepsilon) \neq 0$, will enlarge the second term, resulting in large biases. Thus, the two identifying assumption failures exacerbate each other: having weak instruments compounds problems from confounding or exclusion restriction violations, and vice versa. With invalid instruments, it is possible that the asymptotic bias of the 2SLS estimator is greater than that of the OLS estimator, i.e., $\left|\frac{\text{Cov}(z,\varepsilon)}{\text{Cov}(z,d)}\right| \gg \left|\frac{\text{Cov}(d,\varepsilon)}{\mathbb{V}[d]}\right|$ in the single instrument case.

While the inference problem can be alleviated by employing alternative inferential methods as described above, addressing the failure of Assumption 2 is more challenging since it is fundamentally a research design issue that should be tackled at the design stage. Researchers often devote significant effort to arguing for unconfoundedness and exclusion restrictions in their settings. In Section A3 of the SM, we provide an exposition of the "zero-first-stage" (ZFS) test (Bound and Jaeger, 2000), which is essentially a placebo test on a subsample where the instrument is expected to be uncorrelated with the treatment, to help researchers gauge the validity of their instruments. These estimates can then be used to debias the 2SLS estimate using the methods proposed in Conley, Hansen and Rossi (2012).

3. Data and Types of Instruments

In this section, we first discuss our case selection criteria and the replication sample, which is the focus of our subsequent analysis. We then describe the types of instruments in the replicable studies.

Data. We examine all empirical papers published in the APSR, AJPS, and JOP from 2010 to 2022 and identify studies that use an IV strategy as one of the main identification strategies, including papers that use binary or continuous treatments and that use a single or multiple instruments. We use the following criteria: (1) the discussion of the IV result needs to appear in the main text and support a main argument in the paper; (2) we consider linear models only; in other words, papers that use discrete outcome models are excluded from our sample; (3) we exclude papers that include multiple endogenous variables in a single specification (multiple endogenous variables in separate specifications are included); (4) we exclude papers that use IV or GMM estimators in a dynamic panel setting because they are subject to a separate set of empirical issues and their poor performance has been thoroughly discussed in the literature (Bun and Windmeijer, 2010). These criteria result in 30 papers in the APSR, 33 papers in the AJPS, and 51 papers in the JOP. We then strive to find replication materials for these papers from public data-sharing platforms, such as the Harvard Dataverse, and the authors' websites. We are able to locate complete replication materials for 76 (62%) papers. However, code completeness and quality of documentation vary a great deal. Data availability has significantly improved since 2016-2017 following new editorial policies requiring authors to make replication materials publicly available, though none of the journals requires full replicability administrated by a third party as a condition for publication (Key, 2016), which would constitute a major improvement in our view.

Using data and code from the replication materials, we set out to replicate the main IV

Table 1. Data availability and replicability of IV papers.

	#All Papers	Incomplete Data	Incomplete Code	Replication Error	Replicable
APSR	30	16	0	3	14 (42%)
AJPS	33	3	1	1	25 (83%)
JOP	51	19	3	1	28 (55%)
Total	114	38	4	5	67 (59%)

results in the 67 papers. Our replicability criterion is simple: As long as we can exactly replicate *one* 2SLS point estimate that appears in the paper, we deem the paper replicable. We do not aim at exactly replicating SEs, z-scores, or level of statistical significance for the 2SLS estimates because they involve the choice of the inferential method, which we will discuss in the next section.

After much effort and hundreds of hours of work, we are able to replicate the main results of 61 papers.[6] The low replication rate is consistent with what is reported in Hainmueller, Mummolo and Xu (2019). The main reasons for failures of replication are incomplete data (38 papers), incomplete code or poor documentation (4 papers), and replication errors (5 papers). Table 1 presents summary statistics on data availability and replicability of IV papers for each of the three journals. The rest of this paper focuses on results based on these 67 replicable papers (and 70 IV designs).

Types of instruments. Inspired by Sovey and Green (2011), in Table 2, we summarize the types of IVs in the replicable designs, although our categories differ from theirs to reflect changes in the types of instruments used in the discipline. As in Sovey and Green (2011), the biggest category is "Theory," in which the authors justify Assumption 2, including IVs' quasi-randomness and the exclusion restriction, using social science theories or substantive knowledge. We further divide theory-based IVs into four subcategories: geogra-

[6]For three papers, we are able to produce the 2SLS estimates with perfectly executable code; however, our replicated estimates are inconsistent with what was reported in the original studies. We suspect the inconsistencies are caused by data rescaling or misreporting; hence, we keep them in the sample.

phy/climate/weather, history, treatment diffusion, and others.

Many studies in the theory category justify the choices of their instruments based on geography, climate, or weather conditions. For example, Zhu (2017) uses weighted geographic closeness as an instrument for the activities of multinational corporations; Hager and Hilbig (2019) use mean elevation and distance to rivers to instrument equitable inheritance customs; and Grossman, Pierskalla and Boswell Dean (2017*a*) use the number of distinct landmasses as an instrument for government fragmentation. Henderson and Brooks (2016) use rainfall around Election Day as an IV for democratic vote margins. The popularity of weather instruments for a whole host of outcomes necessarily implies that the exclusion restriction is especially tenuous in such cases (Mellon, 2020).

TABLE 2. TYPES OF INSTRUMENTS

Type	#Papers	Percentage %
Theory	42	60.0
Geography/climate/weather	13	18.6
History	11	15.7
Treatment diffusion	2	2.9
Others	16	22.9
Experiment	12	17.1
Econometrics	9	12.9
Interactions/"Bartik"	7	10.0
Lagged treatment	1	1.4
Empirical test	1	1.4
Rules & policy changes	7	10.0
Change in exposure	3	4.3
Fuzzy RD	4	5.7
Total	70	100.0

Historical instruments are based on historical differences between units that cannot be explained by current levels of the treatment. For example, Vernby (2013) uses historical immigration levels as an instrument for the current number of non-citizen residents. Similarly, Spenkuch and Tillmann (2018) use historical decisions by rulers in Europe over the religion of their region to instrument for the current religion of survey respondents. These studies use historical variation as instruments for current or modern variables.

Several studies base their choices on regional diffusion of treatment. For example, Dube and Naidu (2015) use US military aid to countries outside Latin America as an instrument for US military aid to Colombia. Grossman, Pierskalla and Boswell Dean (2017*b*) use over-time variation in the number of regional governments to instrument government fragmentation in sub-Saharan Africa. Dorsch and Maarek (2019) use the regional share of democracies as an instrument for democratization in a country-year panel.

Finally, several papers rely on a unique instrument based on theories that we could not place in a category. For example, Carnegie and Marinov (2017) use the rotating presidency of the Council of the European Union as an instrument for official development aid. They argue that countries that were colonized by the country that holds the presidency receive exogenously more aid than other countries. Dower et al. (2018) use religious polarization as an instrument for the frequency of unrest and argue that religious polarization could only impact collective action through its impact on representation in local institutions.

The second-biggest category is randomized experiments. Articles in this category employ randomization, designed and conducted by researchers or a third party, to make causal inference and use 2SLS estimation to address non-compliance issues in an encouragement design—the IV normally is being encouraged to take the treatment. With random assignment, we have more confidence in Assumption 2 because $z \perp\!\!\!\perp v$ by design, and the direct effect of encouragement on the outcome is easier to rule out than without random assignment.

Another category of instruments are based on explicit rules, which generate quasi-random variation in the treatment. Sovey and Green (2011) refer to this category as "Natural Experiment." We avoid this terminology because it is widely misused. We limit this category to two circumstances: fuzzy regression discontinuity (RD) designs and variation in exposure to policies due to time of birth or eligibility.[7] For example, Kim (2019) leverages a reform in

[7]The difference between the two is subtle: For the latter, the gap in the forcing variable, such as birth

Sweden that requires municipalities above a population threshold to adopt direct democratic institutions. Dinas (2014) uses eligibility to vote based on age at the time of an election as an instrument for whether respondents did vote.

The last category of instruments are based on econometric assumptions. This category includes what Sovey and Green (2011) call "Lags." These are econometric transformations of variables argued to constitute instruments. For example, Lorentzen, Landry and Yasuda (2014) use a measure of the independent variable from 8 years earlier to mitigate endogeneity concerns. Another example is instruments relying on variable transformations to satisfy assumptions, such as Shift-share "Bartik" instruments based on interactions between multiple variables. For example, Baccini and Weymouth (2021) use the share of jobs in a specific industry within a county, interacted with national-level changes in employment in that industry, to study the effect of manufacturing layoffs on voting.

Compared to IV papers published before 2010, there is a significant increase in the proportion of papers using experiment-generated IVs (from 2.9% to 17.1%) due to the growing popularity of survey and field experiments. In contrast, the number of papers relying on econometric techniques or flawed empirical tests (such as regressing y on d and z and checking if the coefficient of z is significant) has decreased, thanks to improving empirical practices in the discipline. The percentage of papers using theory-justified instruments remains nearly the same at around 60%.

4. Replication Procedure and Results

In this section, we describe our replication procedure and report the main findings.

Procedure. For each paper, we select the main IV specification that plays a central role in supporting a main claim in the paper; it is either referred to as the baseline specification

cohort, is fixed and cannot be arbitrarily small.

or appears in one of the main tables or figures. Focusing on this specification, our replication procedure involves the following steps. First, we compute the first-stage partial F statistics based on (1) classic analytic SEs, (2) Huber White heteroskedastic-robust SEs, (3) cluster-robust SEs (if applicable and based on the original specifications), and (4) bootstrapped SEs.[8] We also calculate $F_{\texttt{Eff}}$.

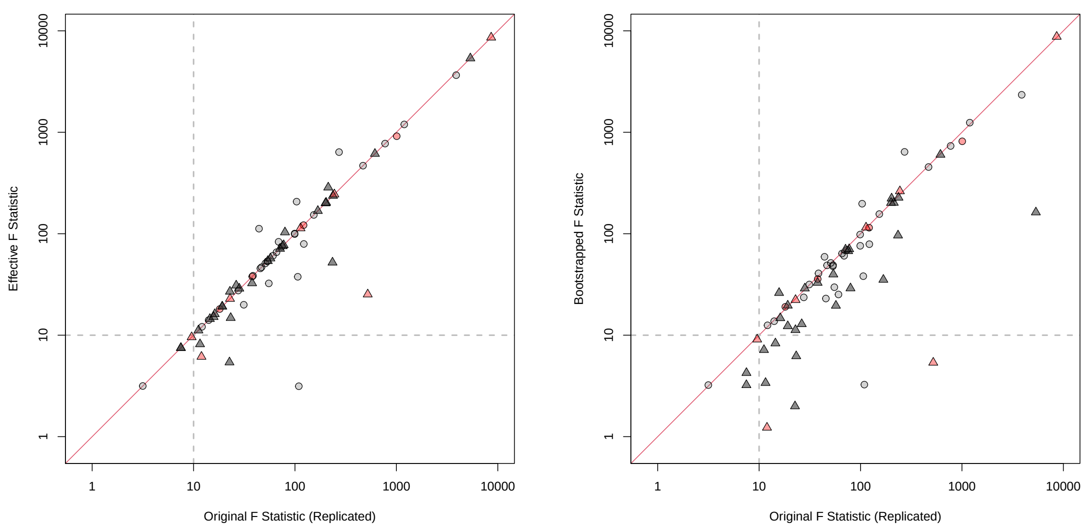
We then replicate the original IV result using the 2SLS estimator and apply four different inferential procedures. First, we make inferences based on analytic SEs, including robust SEs or cluster-robust SEs (if applicable). Additionally, we use two nonparametric bootstrap procedures, as described in Section 2, *bootstrap-c* and *bootstrap-t*. For specifications with only a single instrument, we also employ the tF procedure proposed by Lee et al. (2022), using 2SLS t-statistics and first-stage F-statistics based on analytic SEs accounting for the originally specified clustering structure. Finally, we conduct an AR procedure and record the p-values and CIs.

We record the point estimates, SEs (if applicable), 95% CIs, and p-values for each procedure (the point estimates fully replicate the reported estimates in the original papers and are the same across all procedures). In addition, we estimate a naïve OLS model by regressing the outcome variable on the treatment and control variables, leaving out the instrument. We calculate the ratio between the magnitudes of the 2SLS and OLS estimates. We also record other useful information, such as the number of observations, the number of clusters, the types of instruments, the methods used to calculate SEs or CIs, and the rationale for each paper's IV strategy. Our replication yields the following three main findings.

[8]They are calculated by $F_{boot} = \hat{\tau}_{2SLS}' \hat{\mathrm{Var}}_{boot}(\hat{\tau}_{2SLS})^{-1} \hat{\tau}_{2SLS}/p_z$, where p_z is the number of IVs and $\hat{\mathrm{Var}}_{boot}(\hat{\tau}_{2SLS})$ is the estimated variance-covariance matrix based on a nonparametric bootstrap procedure, in which we repeatedly sample the rows of the data matrix with replacement. If the data have a clustered structure, we use cluster-bootstrapping instead by sampling with replacement each cluster of data (Colin Cameron and Miller, 2015; Esarey and Menger, 2019). We include F_{boot} as a reference to the classic F and effective F. In Section A.2 of the SM, we compare the five types of F statistics and show that the effective F and F based on bootstrapping are usually more conservative than other F statistics.

Finding 1. First-stage partial F statistics. Our first finding regards the strengths of the instruments. To our surprise, among the 70 IV designs, 12 (17%) do not report this crucial statistic despite its key role in justifying the validity of an IV design. Among the remaining 58 studies that report F statistics, 9 (16%) use classic analytic SEs, thus not adjusting for potential heteroskedasticity or clustering structure. In Figure 3, we plot the replicated first-stage partial F statistics based on the authors' original model specifications and choices of variance estimators on the x-axis against effective F statistics (a) or bootstrapped F statistics (b) on the y-axis. Both axes are on a logarithmic scale.[9]

FIGURE 3. ORIGINAL VS. EFFECTIVE AND BOOTSTRAPPED F



(a) Original F vs. Effective F (b) Original F vs. Bootstrapped F

Note: Circles and triangles represent applications with and without a clustering structure, respectively. Studies that do not report F statistics are painted in red. The original F statistics are replicated based on the authors' original model specifications and choices of variance estimators in the 2SLS regressions. They may differ from those reported in the papers because of misreporting.

In the original studies, the authors used various SE estimators, such as classic SEs, robust SEs, or cluster-robust SEs. As a result, the effective F may be larger or smaller than the original ones. However, a notable feature of Figure 3 is that when a clustering structure exists, the original F statistics tend to be larger than the effective F or bootstrapped F.

[9]We use the replicated F statistics instead of the reported ones because some authors either do not report or misreport their F statistics (see SM for a comparison between the reported and replicated F statistics).

When using the effective F as the benchmark, 8 studies (11%) have $F_{\texttt{Eff}} < 10$. This number increases to 12 (17%) when the bootstrapped F statistics are used. The median first-stage $F_{\texttt{Eff}}$ statistic is higher in experimental studies compared to non-experimental ones (67.7 versus 53.5). It is well known that failing to cluster the SEs at appropriate levels or using the analytic cluster-robust SE with too few clusters can lead to a severe overstatement of statistical significance (Cameron, Gelbach and Miller, 2008). However, this problem has received less attention when evaluating IV strength using F statistics.

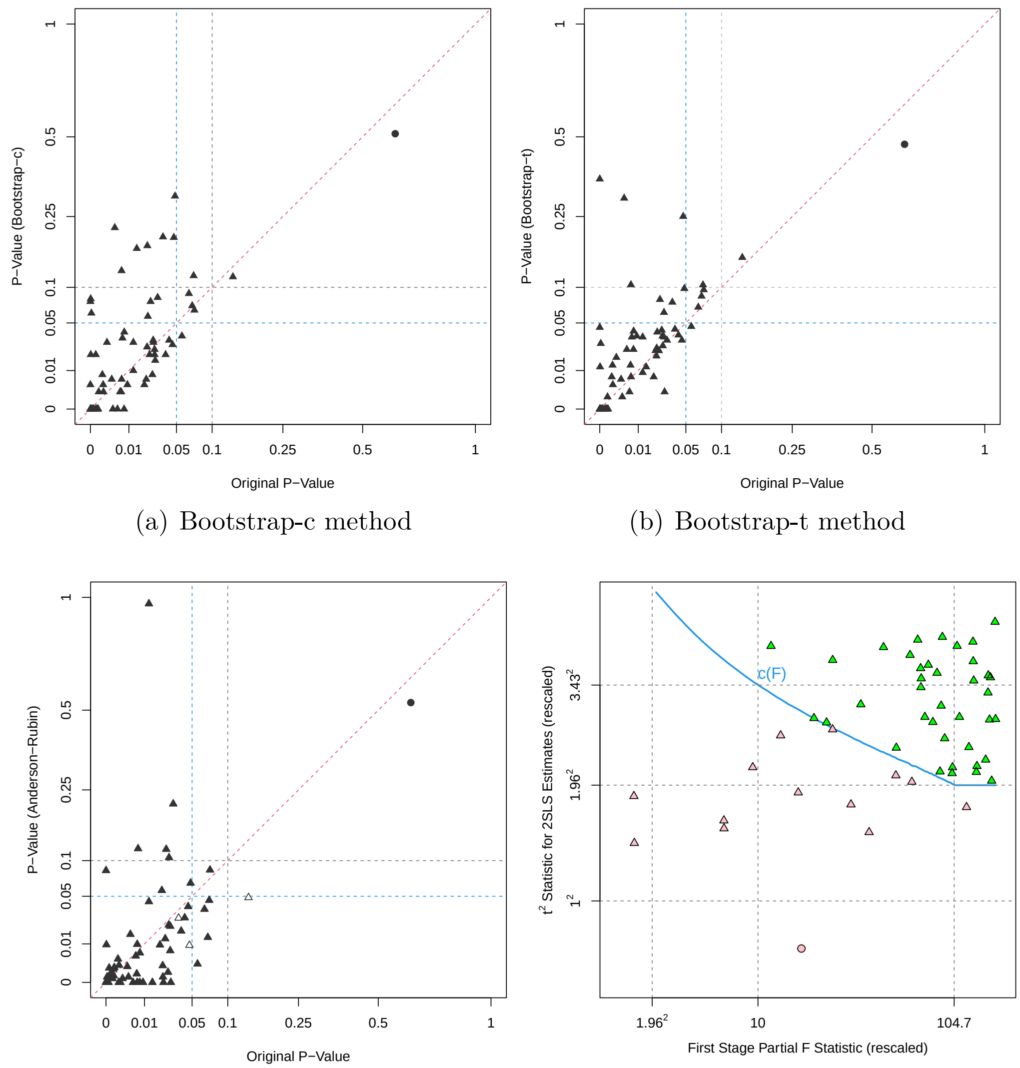
Finding 2. Inference. Next, we compare the reported and replicated p-values for the null hypothesis of no effect. For studies that do not report a p-value, we calculate it based on a standard normal distribution using the reported point estimates and SEs. The replicated p-values are based on (1) *bootstrap-c*, (2) *bootstrap-t*, and (3) the AR procedure. Since we can exactly replicate the point estimates for the papers in the replication sample, the differences in p-values are the result of the inferential methods used. Figure 4(a)-(c) plot reported and replicated p-values, from which we observed two patterns. First, most of the reported p-values are smaller than 0.05 or 0.10, the conventional thresholds for statistical significance. Second, consistent with Young (2022)'s finding, our replicated p-values based on the AR procedure or bootstrap methods are usually bigger than the reported p-value (exceptions are mostly caused by rounding errors), which are primarily based on t statistics calculated using analytic SEs. Using the AR test, we cannot reject the null hypothesis of no effect at the 5% level in 13 studies (19%), compared with 7 (10%) in the original studies. The number increases to 15 (21%) and 20 (29%) when we use p-values from the *bootstrap-t* and *-c* methods. Note that very few papers we review utilize inferential procedures specifically designed for weak instruments, such as the AR test (2 papers), the conditional likelihood-ratio test (Moreira, 2003) (1 paper), and confident sets (Mikusheva and Poi, 2006) (none).

We also apply the tF procedure to 54 studies that use single IVs using $F_{\texttt{Eff}}$ statistics and t statistics based on robust or cluster-robust SEs. Figures 4(d) shows that 16 studies (30%) are not statistically significant at the 5% level, and 5 studies deemed statistically significant when using the conventional fixed critical values for the t-test become statistically insignificant using the tF procedure, indicating that overly optimistic critical values due to weak instruments also contribute to overestimation of statistical power, but not as the primary factor. These results suggest that both weak instruments and non-i.i.d. errors have contributed to severe overstatements of power in IV studies in political science.

Finding 3. 2SLS-OLS discrepancy. Finally, we investigate the relationship between the 2SLS estimates and naïve OLS estimates. In Figure 5(a), we plot the 2SLS coefficients against the OLS coefficients, both normalized using reported OLS SEs. The shaded area indicates the range beyond which the OLS estimates are statistically significant at the 5% level. It shows that for most studies in our sample, the 2SLS estimates and OLS estimates share the same direction and that the magnitudes of the former are often much larger than those of the latter. Figure 5(b) plots the distribution of the ratio between the 2SLS and OLS estimates (in absolute terms). The mean and median of the absolute ratios are 12.4 and 3.4, respectively. In fact, in all but two designs (97%), the 2SLS estimates are bigger than the OLS estimates, consistent with Jiang (2017)'s finding based on finance research. While it is theoretically possible for most OLS estimates in our sample to be biased towards zero, only 21% of the studies have researchers expressing their belief in downward biases of the OLS estimates. Meanwhile, 40% of the studies consider the OLS results to be their main findings. The fact that researchers use IV designs as robustness checks for OLS estimates due to concerns of upward biases is apparently at odds with the significantly larger magnitudes of the 2SLS estimates.

In Figure 5(c), we further explore whether the 2SLS-OLS discrepancy is related to IV strength, measured by $\hat{\rho}(d, \hat{d})$, the estimated correlation coefficient between the treatment

FIGURE 4. ALTERNATIVE INFERENTIAL METHODS.



(a) Bootstrap-c method

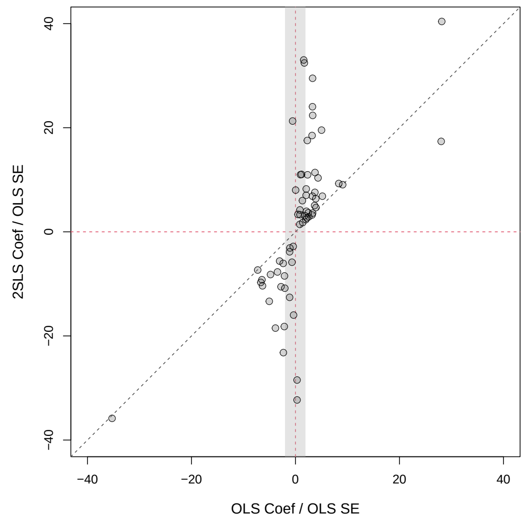
(b) Bootstrap-t method

(c) Anderson-Rubin

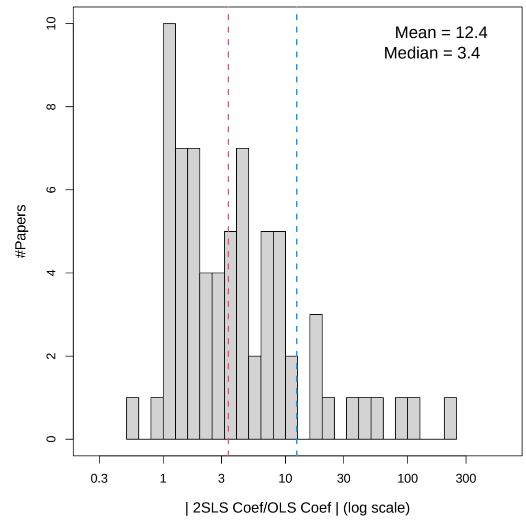
(d) tF procedure

Note: In subfigures (a)-(c), we compare original p-values to those from alternative inferential methods, testing against the null that $\tau = 0$. Both axes use a square-root scale. Original p-values are adapted from original papers or calculated using standard-normal approximations of z-scores. Solid circles represent Arias and Stasavage (2019), where authors argue for a null effect using IV strategy. *Bootstrap-c* and *-t* represent percentile methods based on 2SLS estimates and t-statistics, respectively, using original model specifications. Hollow triangles in subfigure (c) indicate unbounded 95% CIs from the AR test using the inversion method. Subfigure (d) presents tF procedure results from 54 single instrument designs. Green and red dots represent studies remaining statistically significant at the 5% level using the tF procedure and those that don't, respectively. Subfigures (a)-(c) are inspired by Figure 3 in Young (2022), and subfigure (d) by Figure 3 in Lee et al. (2022).

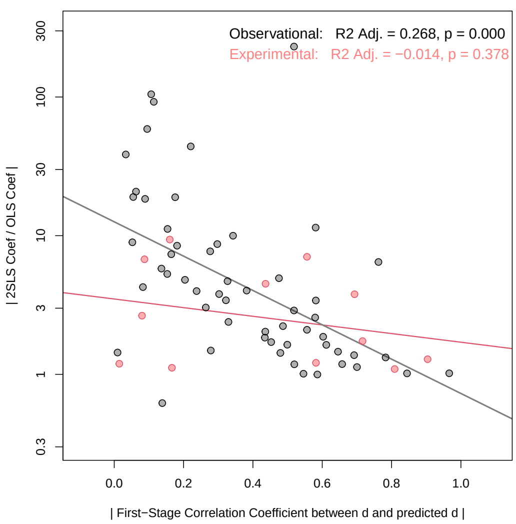
Figure 5. Relationship between OLS and 2SLS Estimates



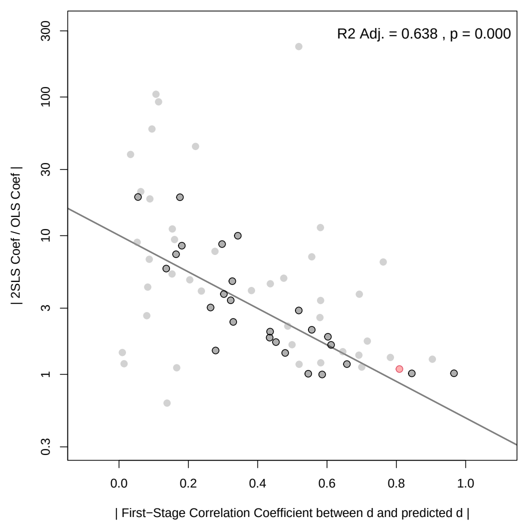
(a) OLS vs 2SLS Estimates



(b) Distribution of Ratio



(c) First Stage and the Ratio: Full Sample



(d) First Stage and Ratio: Subsample

Note: Subfigures (a) and (b) use reported 2SLS and OLS coefficient estimates. In subfigure (a), both axes are normalized by reported OLS SE estimates with the gray band representing the $[-1.96, 1.96]$ interval. Subfigures (c) and (d) feature the relationship between the correlational coefficient between d and $\hat{d}$ and the ratio of 2SLS and OLS estimates. Gray and red circles represent observational and experimental studies, respectively. Subfigure (d) highlights studies with statistically significant OLS results at the 5% level, claimed as part of the main findings.

and predicted treatment. We find a strong negative correlation between $|\hat{\tau}_{2SLS}/\hat{\tau}_{OLS}|$ and $|\hat{\rho}(d, \hat{d})|$ among studies using non-experimental instruments (grey dots). The adjusted R^2

is 0.268, with $p = 0.000$. However, the relationship is much weaker among studies using experiment-generated instruments (red dots). The adjusted R^2 is -0.014 with $p = 0.378$. In Figure 5(d), we limit our focus to the subsample in which the OLS estimates are statistically significant at the 5% level and researchers accept them as (part of) the main findings, and the strong negative correlation remains. At first glance, this result may seem mechanical: as the correlation between d and $\hat{d}$ increases, the 2SLS estimates naturally converge to the OLS estimates. However, the properties of the 2SLS estimator under the identifying assumptions do not predict the negative relationship (we confirm it in simulations in the SM), and such a relationship is not found in experimental studies.

We believe that several factors contribute to this pattern, including (1) the failure of Assumption 2, (2) publication bias, (3) HTE, and (4) measurement error in d. We suspect the first two factors are the main driving forces. As previously mentioned, when Assumption 2 is violated, weak instruments amplify the biases from endogenous IVs or exclusion restriction failures, i.e., $\frac{\text{Bias}_{IV}}{\text{Bias}_{OLS}} = \frac{\text{Cov}(z,\varepsilon)\mathbb{V}[d]}{\text{Cov}(z,d)\text{Cov}(d,\varepsilon)} \gg 1$. Publication bias may also play a role. When the first stage is weak, IV estimates have a larger variance and can be very large or very small in magnitude compared to OLS estimates. If researchers selectively report statistically significant results or journals tend to publish papers with statistically significant findings, we may observe a negative relationship as in Figure 5. This phenomenon is also referred to as Type-M bias in the psychology and sociology literature (Gelman and Carlin, 2014; Felton and Stewart, 2022).

Moreover, 30% of the replicated studies in our sample mention HTE as a possible explanation for this discrepancy. OLS and 2SLS place different weights on covariate strata in the sample, and therefore if compliers, those whose treatment status is affected by the instrument, are more responsive to the treatment than the rest of the units in the sample, we might see diverging OLS and 2SLS estimates. Under the assumption that the exclusion restriction holds, this gap can be decomposed into covariate weight difference, treatment-level weight difference, and endogeneity bias components using the procedure developed in (Ishimaru, 2021). In the SM, we investigate this possibility and find that it is highly unlikely that HTE *alone* can explain the difference in magnitudes between 2SLS and OLS estimates we observe in the replication data, i.e., the variance in treatment effects needed for this gap is implausibly large.

Finally, an IV design can correct for the downward bias of the measurement error in d, resulting in $|\hat{\tau}_{2SLS}/\hat{\tau}_{OLS}| > 1$. If the measurement error is large, this can weaken the relationship between d and $\hat{d}$, producing a negative correlation. However, it is worth noting that only 4 papers in our sample (6%) attribute the IV strategy to measurement error; the negative correlation remains even when the OLS estimates are the main findings (indicating measurement error may not be as concerning for researchers).

We summarize the main findings from our replication exercise in Table 3. The three issues we have identified are observed in all three journals included in the study. Based on these results, we believe that a significant portion of the IV results either lack credibility or do not provide new information beyond what is already provided by OLS regressions.

5. Recommendations

IV designs in experimental and observational studies differ fundamentally. In randomized experiments, the instruments' unconfoundedness is ensured by design, and researchers can address potential exclusion restriction failures at the design stage, e.g., by testing potential design effects through randomization (Gerber and Green, 2012, pp. 140-141). Practices like power analysis, placebo tests, and preregistration in experimental studies also help reduce improper use of IV designs. In contrast, analyzing observational IV design based on "natural experiments" requires detailed knowledge of the assignment mechanism, making them more complex and prone to potential issues (Sekhon and Titiunik, 2012).

Our findings suggest that using an IV strategy in an observational setting is much more

Table 3. Summary of Replication Results

(%)	APSR (15)	AJPS (25)	JOP (30)	All (70)
Panel A: First-Stage F Statistic				
Unreported	0.0	20.0	23.3	17.1
Reported $F > 1.3$ effective F	20.0	25.0	30.4	25.9
Effective $F < 10$	13.3	12.0	10.0	11.4
Bootstrapped $F < 10$	13.3	20.0	16.7	17.1
Panel B: Inference for IV Designs				
Original $p > 0.05$	20.0	8.0	6.7	10.0
AR $p > 0.05$	13.3	24.0	16.7	18.6
Bootstrap-c $p > 0.05$	20.0	32.0	30.0	28.6
Bootstrap-t $p > 0.05$	26.7	24.0	16.7	21.4
tF procedure $p > 0.05$	38.5	23.6	29.2	29.6
Panel C: 2SLS-OLS Relationship				
$sign(\hat{\tau}_{2SLS}) = sign(\hat{\tau}_{OLS})$	93.3	100.0	86.7	92.9
$\|\hat{\tau}_{2SLS}/\hat{\tau}_{OLS}\| > 1$	93.3	100.0	96.7	97.1
$\|\hat{\tau}_{2SLS}/\hat{\tau}_{OLS}\| > 3$	53.3	44.0	60.0	52.9
$\|\hat{\tau}_{2SLS}/\hat{\tau}_{OLS}\| > 5$	40.0	32.0	33.3	34.3
$\|\hat{\tau}_{2SLS}/\hat{\tau}_{OLS}\| > 10$	13.3	16.0	20.0	17.1

challenging. Since unconfoundedness is not guaranteed by design, researchers have a greater burden of proof for the validity of IVs. On the one hand, truly random (and strong) instruments are rare; on the other hand, it is difficult to conduct placebo tests, such as the ZFS test, for the exclusion restriction after data collection. Additionally, researchers often cannot easily increase the sample size to obtain sufficient statistical power. To prevent misusing IVs in observational studies, we provide a checklist for researchers to consider when applying or considering applying an IV strategy with observational data (in the case of one endogenous treatment variable):

Design

- Prior to using an IV strategy, consider how selection bias may be affecting treatment effect estimates obtained through OLS. If the main concern is underestimating an already statistically significant treatment effect, an IV strategy may not be necessary.

- During the research design phase, consider whether the chosen instrument can realistically create random or quasi-random variations in treatment assignment while remaining excluded from the outcome equation.

Characterizing the first-stage

- Calculate and report $F_{\texttt{Eff}}$ for the first stage, taking into account heteroscedasticity and clustering structure as needed. However, do not discard a design simply because $F_{\texttt{Eff}} < 0$.
- If d and z are continuous, plot d against its predicted values $\hat{d}$ (with covariates and fixed effects already partialled out from both) and visually verify whether their relationship aligns with theoretical expectations.

Hypothesis testing and inference

- *Option 1. t-test with F_{Eff} pretesting.* If $F_{\texttt{Eff}} < 10$, choose Options 2 or 3. Utilize conservative methods like *bootstrap-t* and *bootstrap-c* if outliers or group structures are present.
- *Option 2. tF procedure.* For single treatment and instrument cases, adjust t-test critical values based on $F_{\texttt{Eff}}$.
- *Option 3. Direct testing.* Apply weak-instrument-robust procedures, such as the AR test.

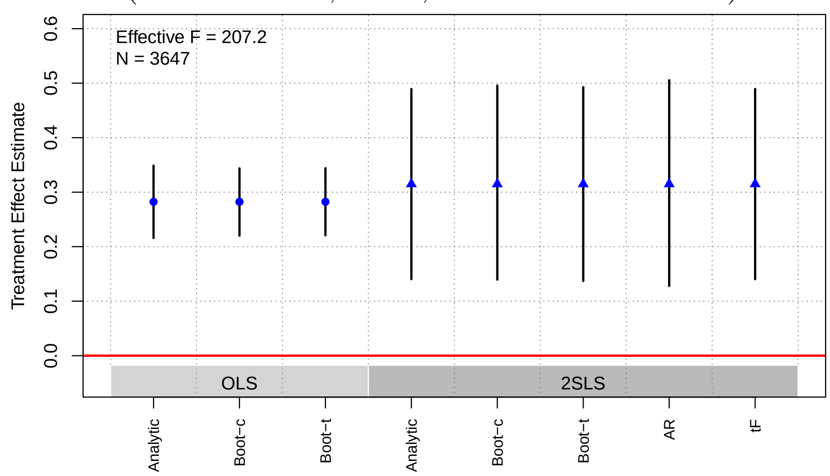
Communicating your findings

- Present OLS and IV estimates alongside CIs from various inferential methods in a graphical format, like in Figure 6. These CIs may not concur on statistical significance, but they collectively convey the findings' robustness to different inferential approaches.
- Remember to report first-stage and reduced-form estimation results, including 95% CIs for coefficients, as they offer insight into both instrument strength and statistical power.

Additional diagnostics

- If you expect the OLS results to be upward biased, be concerned if the 2SLS estimator yields much larger estimates.
- If there is good reason to believe that treatment effects on compliers are significantly larger in magnitude than those on non-compliers, explain this through profiling of these principal strata (Abadie, 2003; Marbach and Hangartner, 2020).
- If it is possible to identify an observational analogue of "never takers" or a subset of them, conduct a placebo test by estimating the effect of the instrument on the outcome of interest in this ZFS sample. Using results from the ZFS test, obtain local-to-zero IV estimates and CIs and compare them to the original estimates and CIs. See the SM for a detailed example.

FIGURE 6. REPLICATED OLS AND 2SLS ESTIMATES WITH 95% CIS (MCCLENDON, 2014, TABLE 2 COLUMN 1)



Note: The treatment is reading an email with a promise of social esteem. The instrument is being encouraged to take the treatment. The outcome is attending LGBTQ events. The AR test does not rely on the first-stage F. Similar figures for each of the 70 IV designs are shown in the SM. This plot is made by ivDiag.

We provide an accompanying `R` package, ivDiag, to implement our recommended procedures. Our aim is to address concerns regarding IVs in social science research and improve the quality of estimation and inference, especially for non-experimental IV designs.

References

Abadie, Alberto. 2003. "Semiparametric Instrumental Variable Estimation of Treatment Response Models." *Journal of Econometrics* 113(2):231–263.

Anderson, Theodore W and Herman Rubin. 1949. "Estimation of the parameters of a single equation in a complete system of stochastic equations." *The Annals of mathematical statistics* 20(1):46–63.

Anderson, Theodore W, Naoto Kunitomo and Takamitsu Sawa. 1982. "Evaluation Of The Distribution Function Of The Limited Information Maximum Likelihood Estimator." *Econometrica: Journal of the Econometric Society* pp. 1009–1027.

Andrews, Donald WK and Patrik Guggenberger. 2009. "Validity Of Subsampling And" Plug-in Asymptotic" Inference For Parameters Defined By Moment Inequalities." *Econometric Theory* pp. 669–709.

Andrews, Isaiah, James Stock and Liyang Sun. 2019. "Weak Instruments In Instrumental Variables Regression: Theory And Practice." *Annual Review of Economics* 11:727–753.

Angrist, Joshua D, Guido W Imbens and Donald B Rubin. 1996. "Identification Of Causal Effects Using Instrumental Variables." *Journal of the American statistical Association* 91(434):444–455.

Angrist, Joshua D and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics*. Princeton university press.

Arellano, Manuel. 2002. "Sargan's Intrumental Variables Estimation And The Generalized Method Of Moments." *Journal of Business & Economic Statistics* 20(4):450–459.

Arias, Eric and David Stasavage. 2019. "How Large Are The Political Costs Of Fiscal Austerity?" *The Journal of Politics* 81(4):1517–1522.

Baccini, Leonardo and Stephen Weymouth. 2021. "Gone for good: Deindustrialization, white voter backlash, and US presidential voting." *American Political Science Review* 115(2):550–567.

Bound, John and David A Jaeger. 2000. "Do Compulsory School Attendance Laws Alone Explain The Association Between Quarter Of Birth And Earnings?" *Research in labor economics* 19(4):83–108.

Bound, John, David A Jaeger and Regina M Baker. 1995. "Problems With Instrumental Variables Estimation When The Correlation Between The Instruments And The Endogenous Explanatory VariableIs Weak." *Journal of the American statistical association* 90(430):443–450.

Bun, Maurice JG and Frank Windmeijer. 2010. "The Weak Instrument Problem Of The System GMM Estimator In Dynamic Panel Data Models." *The Econometrics Journal* 13(1):95–126.

Cameron, A Colin, Jonah B Gelbach and Douglas L Miller. 2008. "Bootstrap-based Improvements For Inference With Clustered Errors." *The Review of Economics and Statistics* 90(3):414–427.

Carnegie, Allison and Nikolay Marinov. 2017. "Foreign Aid, Human Rights, And Democracy Promotion: Evidence From A Natural Experiment." *American Journal of Political Science* 61(3):671–683.

Chernozhukov, Victor and Christian Hansen. 2008. "The reduced form: A simple approach to inference with weak instruments." *Economics Letters* 100(1):68–71.

Cinelli, Carlos and Chad Hazlett. 2022. "An omitted variable bias framework for sensitivity analysis of instrumental variables." *Available at SSRN 4217915* .

Colin Cameron, A and Douglas L Miller. 2015. "A Practitioner's Guide To Cluster-robust Inference." *The Journal of Human Resources* 50(2):317–372.

Conley, Timothy G, Christian B Hansen and Peter E Rossi. 2012. "Plausibly Exogenous." *The review of economics and statistics* 94(1):260–272.

Davidson, Russell and James G MacKinnon. 2015. "Bootstrap Tests For Overidentification In Linear Regression Models." *Econometrics* 3(4):825–863.

Dieterle, Steven G and Andy Snell. 2016. "A Simple Diagnostic To Investigate Instrument Validity And Heterogeneous Effects When Using A Single Instrument." *Labour Economics* 42:76–86.

Dinas, Elias. 2014. "Does Choice Bring Loyalty? Electora lParticipation And The Development Of Party Identification." *American Journal of Political Science* 58(2):449–465.

Dorsch, Michael T. and Paul Maarek. 2019. "Democratization And The Conditional Dynamics Of Income Distribution." *American Political Science Review* 113(2):385–404.

Dower, Paul Castañeda, Evgeny Finkel, Scott Gehlbach and Steven Nafziger. 2018. "Collective Action And Representation In Autocracies: Evidence From Russia's Great Reforms." *American Political Science Review* 112(1):125–147.

Dube, Oeindrila and Suresh Naidu. 2015. "Bases, Bullets, And Ballots: The Effect Of Us Military Aid On Political Conflict In Colombia." *The Journal of Politics* 77(1):249–267. Publisher: University of Chicago Press Chicago, IL.

Esarey, J and A Menger. 2019. "Practical And Effective Approaches To Dealing With Clustered Data." *Political Science Research and Methods* 7(3):541–559.

Felton, Chris and Brandon M. Stewart. 2022. "Handle With Care: A Sociologist's Guide To Causal Inference With Instrumental Variables." Mimeo, Princeton University.

Fieller, Edgar C. 1954. "Some Problems In Interval Estimation." *Journal of the Royal Statistical Society: Series B (Methodological)* 16(2):175–185.

Gelman, Andrew and John Carlin. 2014. "Beyond Power Calculations: Assessing Type S (sign) And Type M (magnitude) Errors." *Perspectives on Psychological Science* 9(6):641–651.

Gerber, Alan S and Donald P Green. 2012. *Field Experiments: Design, Analysis And Interpretation*. New York: W. W. Northon.

Greene, William H. 2003. *Econometric Analysis*. Pearson Education India.

Grossman, Guy, Jan H. Pierskalla and Emma Boswell Dean. 2017*a*. "Government Fragmentation And Public Goods Provision." *The Journal of Politics* 79(3):823–840.

Grossman, Guy, Jan H. Pierskalla and Emma Boswell Dean. 2017*b*. "Government Fragmentation And Public Goods Provision." *The Journal of Politics* 79(3):823–840. Publisher: University of Chicago Press Chicago, IL.

Hager, Anselm and Hanno Hilbig. 2019. "Do Inheritance Customs Affect Political And Social Inequality?" *American Journal of Political Science* 63(4):758–773. Publisher: Wiley Online Library.

Hahn, Jinyong and Zhipeng Liao. 2021. "Bootstrap Standard Error Estimates And Inference." *Econometrica: journal of the Econometric Society* 89(4):1963–1977.

Hainmueller, Jens, Jonathan Mummolo and Yiqing Xu. 2019. "How Much Should We Trust Estimates From Multiplicative Interaction Models? Simple Tools To Improve Empirical Practice." *Political Analysis* .

Hall, Peter and Joel L Horowitz. 1996. "Bootstrap critical values for tests based on generalized-method-of-moments estimators." *Econometrica: Journal of the Econometric Society* pp. 891–916.

Hansen, Lars Peter. 1982. "Large Sample Properties Of Generalized Method Of Moments Estimators." *Econometrica: Journal of the Econometric Society* pp. 1029–1054.

Henderson, John and John Brooks. 2016. "Mediating The Electoral Connection: The Information Effects Of Voter Signals On Legislative Behavior." *The Journal of Politics* 78(3):653–669.

Hirano, Keisuke and Jack R Porter. 2015. "Location properties of point estimators in linear instrumental variables and related models." *Econometric Reviews* 34(6-10):720–733.

Ishimaru, Shoya. 2021. "Empirical Decomposition Of The IV-OLS Gap With Heterogeneous And Nonlinear Effects.".
URL: *http://arxiv.org/abs/2101.04346*

Jiang, Wei. 2017. "Have Instrumental Variables Brought Us Closer To The Truth." *The Review of Corporate Finance Studies* 6(2):127–140.

Key, Ellen M. 2016. "How Are We Doing? Data Access And Replication In Political Science." *PS: Political Science & Politics* 49(2):268–272.

Kim, Jeong Hyun. 2019. "Direct Democracy And Women's Political Engagement." *American Journal of Political Science* 63(3):594–610. Publisher: Wiley Online Library.

Lee, David S, Justin McCrary, Marcelo J Moreira and Jack Porter. 2022. "Valid t-ratio Inference for IV." *American Economic Review* 112(10):3260–90.

Lorentzen, Peter, Pierre Landry and John Yasuda. 2014. "Undermining Authoritarian Innovation: The Power Of China's Industrial Giants." *The Journal of Politics* 76(1):182–194.

Marbach, Moritz and Dominik Hangartner. 2020. "Profiling Compliers And Noncompliers For Instrumental-Variable Analysis." *Political analysis: an annual publication of the Methodology Section of the American Political Science Association* 28(3):435–444.

McClendon, Gwyneth H. 2014. "Social Esteem And Participation In Contentious Politics: A Field Experiment At An LGBT Pride Rally." *American Journal of Political Science* 58(2):279–290. Publisher: Wiley Online Library.

Mellon, Jonathan. 2020. "Rain, Rain, GoAway: 137 Potential Exclusion-restriction Violations For Studies Using Weather As An Instrumental Variable." *Available at SSRN* .

Mikusheva, Anna and Brian P Poi. 2006. "Tests And Confidence Sets With Correct Size When Instruments Are Potentially Weak." *The Stata journal* 6(3):335–347.

Moreira, Marcelo J. 2003. "A Conditional Likelihood Ratio Test For Structural Models." *Econometrica: journal of the Econometric Society* 71(4):1027–1048.

Moreira, Marcelo J. 2009. "Tests with correct size when instruments can be arbitrarily weak." *Journal of Econometrics* 152(2):131–140.

Nelson, Charles and Richard Starz. 1990. "Some Further Results On The Exact Small Sample Properties Of The Instrumental Variables Estimator." *Econometrica* 58(41):967–976.

Olea, José Luis Montiel and Carolin Pflueger. 2013. "A Robust Test For Weak Instruments." *Journal of business & economic statistics: a publication of the American Statistical Association* 31(3):358–369.

Sekhon, Jasjeet S and Rocio Titiunik. 2012. "When Natural Experiments Are Neither Natural Nor Experiments." *American Political Science Review* pp. 35–57.

Sovey, Allison J and Donald P Green. 2011. "Instrumental Variables Estimation In Political Science :A Readers' Guide." *American Journal of Political Science* 55(1):188–200.

Spenkuch, Jorg L. and Philipp Tillmann. 2018. "Elite Influence? Religion And The Electoral Success Of The Nazis." *American Journal of Political Science* 62(1):19–36.

Staiger, Douglas and James H Stock. 1997. "Instrumental Variables Regression With Weak Instruments." *Econometrica: journal of the Econometric Society* 65(3):557–586.

Stock, James and Motohiro Yogo. 2005. "Asymptotic Distributions Of Instrumental Variables Statistics With Many Instruments." *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg* pp. 109–120.

Vernby, Kare. 2013. "Inclusion And Public Policy: Evidence From Sweden's Introduction Of Noncitizen Suffrage." *American Journal of Political Science* 57(1):15–29.

Young, Alwyn. 2022. "Consistency Without Inference: Instrumental Variables In Practical Application." *European Economic Review* 147.

Zhu, Boliang. 2017. "MNCs, Rents, And Corruption: Evidence From China." *American Journal of Political Science* 61(1):84–99.

A. Supplementary Materials – Appendix A

Table of Contents

A.1. Additional Information on the Replication Sample

A.1.1. Replication Sample

Figure A1 plots the histograms of effective F statistics using experiment-generated IVs (dark gray) and non-experimental IVs (light gray). The median effective F for experimental and observational designs are 67.7 and 53.5, respectively.

FIGURE A1. HISTOGRAM OF EFFECTIVE F STATISTIC

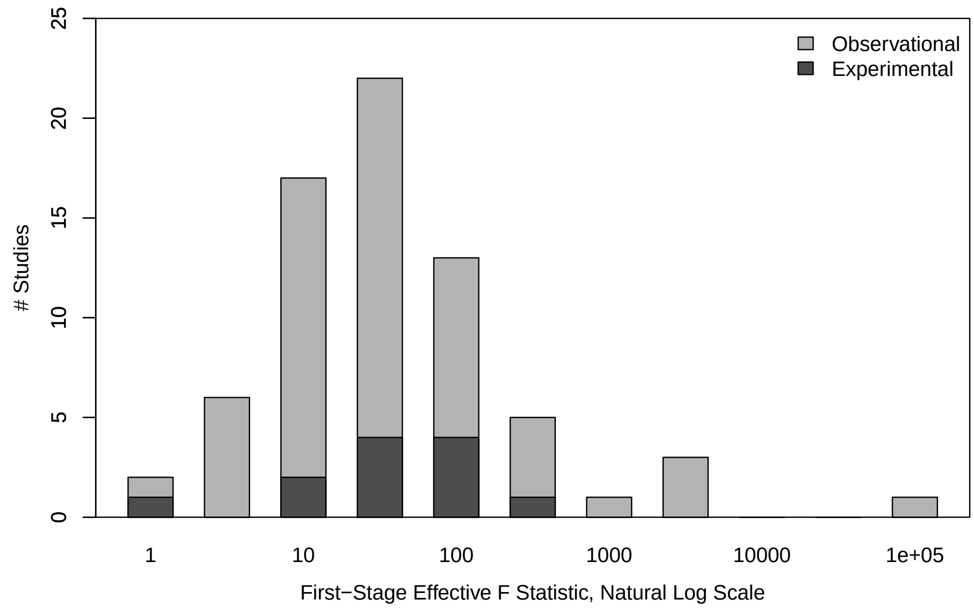
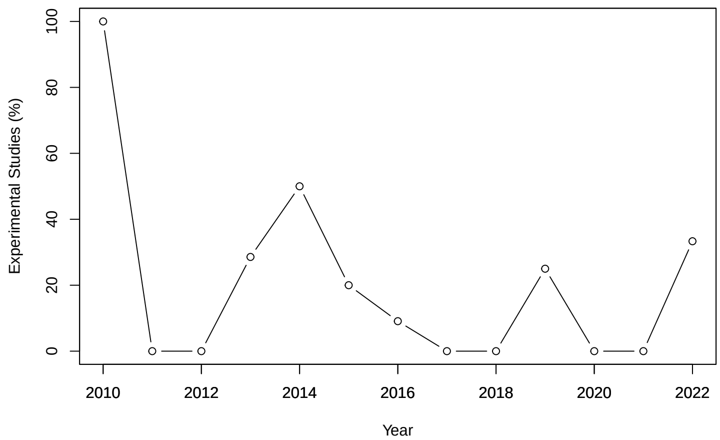
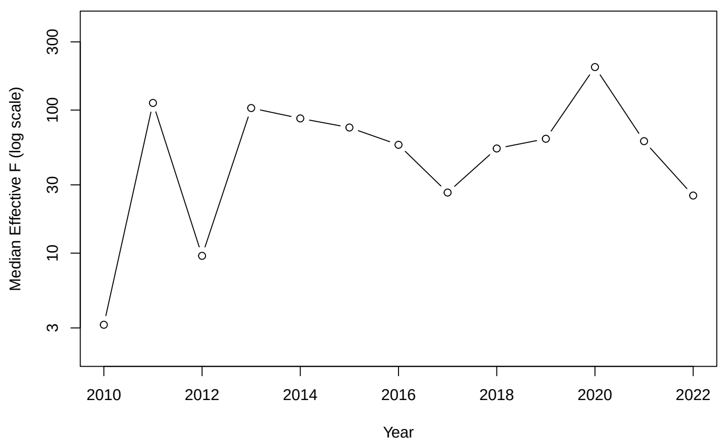


Figure A2 shows the overtime change in the percentage of experimental studies (a), the median effective F statistics (b), and the median ratio between 2SLS and OLS coefficients (c) in the replication sample.

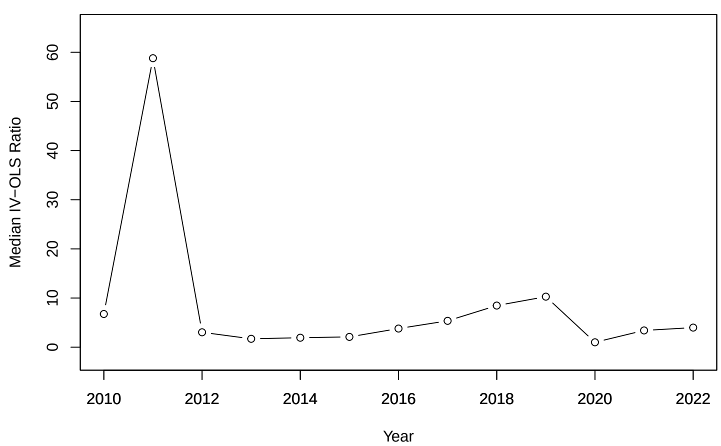
Figure A2. Additional Information on the Sample



(a) Percentage of experimental studies



(b) Effective F statistics (median)

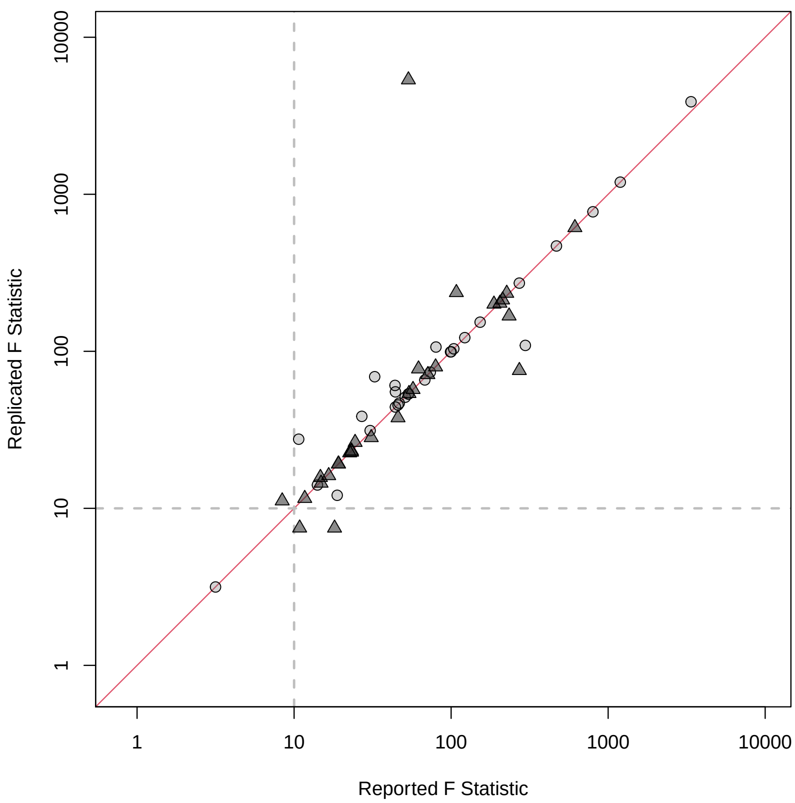


(c) Ratio between 2SLS and OLS coefficients

A.1.2. Comparison of Multiple F Statistics

Figure A3 compares the reported and replicated first-stage partial F statistics (for studies that have reported the F statistics). The replicated F statistics are based on the authors' chosen model specifications and variance estimators in 2SLS estimation. The discrepancy arises from the fact that some authors report the first-stage F statistic based on a different variance estimator than the one used in the 2SLS estimation. In the paper, we use the replicated ones to maintain consistency.

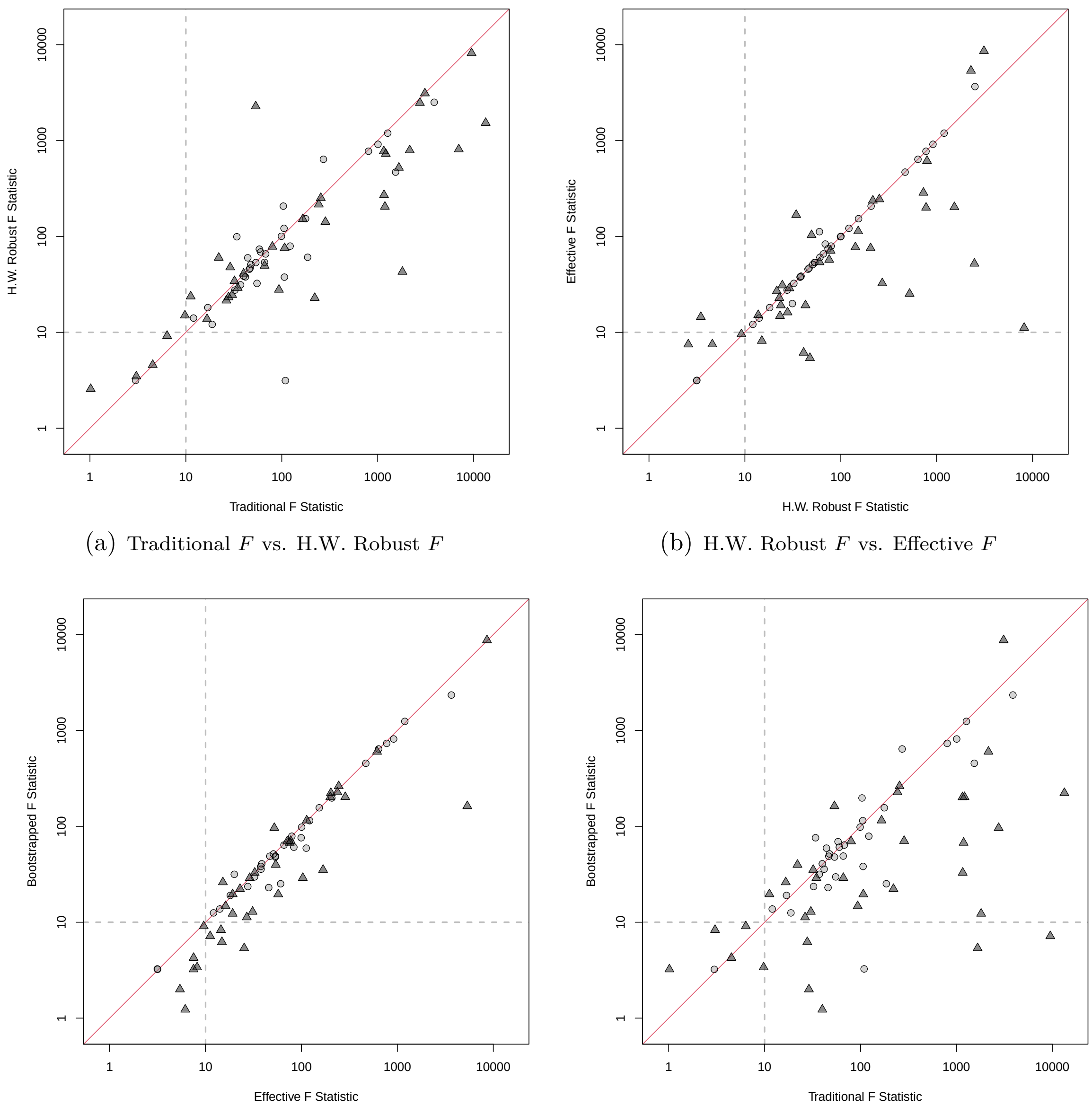
FIGURE A3. REPORTED VS. REPLICATED F STATISTICS



Note: Circles and triangles represent applications with and without a clustering structure, respectively. Studies that do not report F statistics are not shown.

In Figure A4, we compare the traditional F statistics (based on classic analytic SEs), the Huber White robust F statistics, the effective F statistics (robust or cluster-bootstrap SEs) and (cluster-)bootstrapped F statistics. It shows that (cluster-)bootstrapped F statistics are usually the most conservative (smallest). Circles and triangles represent applications with and without a clustering structure, respectively.

FIGURE A4. COMPARISON OF DIFFERENT F STATISTICS



(a) Traditional F vs. H.W. Robust F

(b) H.W. Robust F vs. Effective F

(c) Effective F vs. Bootstrapped F

(d) Traditional F vs. Bootstrapped F

Note: Circles and triangles represent applications with and without a clustering structure, respectively.

A.2. Monte Carlo Evidence

A.2.1. Comparing F Tests for Detecting Weak Instruments

We conduct a simulation study with a clustered DGP in order to evaluate the relative performance of analytic and bootstrap F tests to detect weak instruments. We simulate data from the following DGP

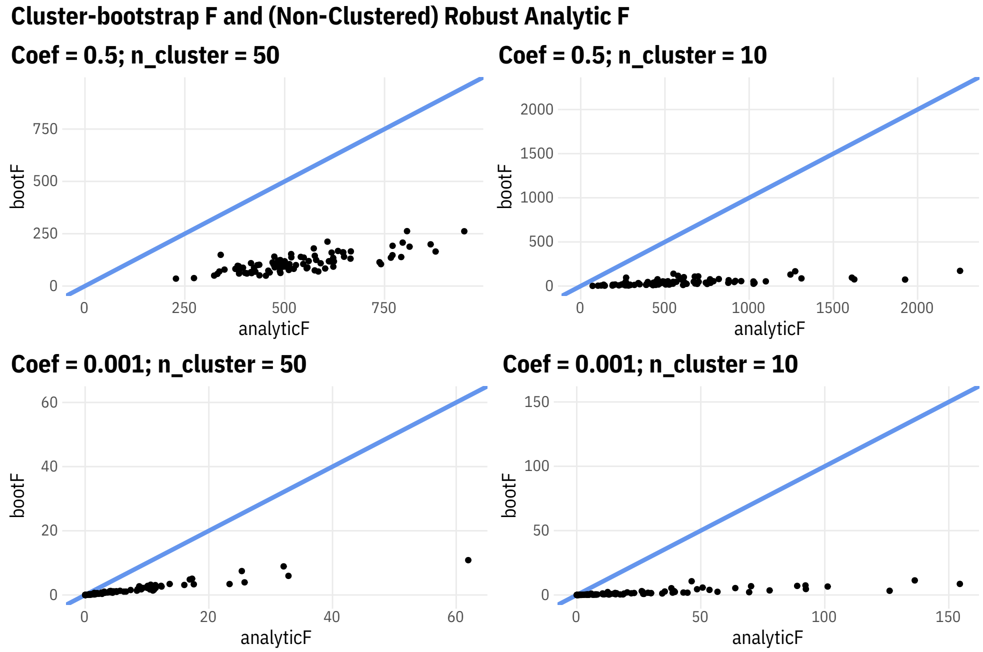
$$\begin{aligned} \text{clustered instrument and error components} \quad \nu_j, \eta_j &\sim \mathcal{N}(0, 0.5) \\ \text{instrument} \quad z_i &\sim \mathcal{N}(0,1) + \nu_j \\ \text{error} \quad \varepsilon_i &\sim \mathcal{N}(0,1) + \eta_j \\ \text{endogenous variable} \quad x_i &= \pi z_i + \varepsilon_i \end{aligned}$$

with errors and instrument components drawn from J clusters. This DGP ensures that the data has dependent structure within each cluster j. We then evaluate the strength of the instrument analytically by computing the t-statistic for $H_0 : \pi = 0$, or by using the corresponding bootstrap statistic $\frac{\pi^2}{\widehat{\sigma}^2}$ where $\widehat{\sigma}^2$ is the bootstrap estimate of the variance of π. We evaluate the analytic and bootstrap F statistics for various values of π and J for 100 replications of the above DGP in Figure (A5).

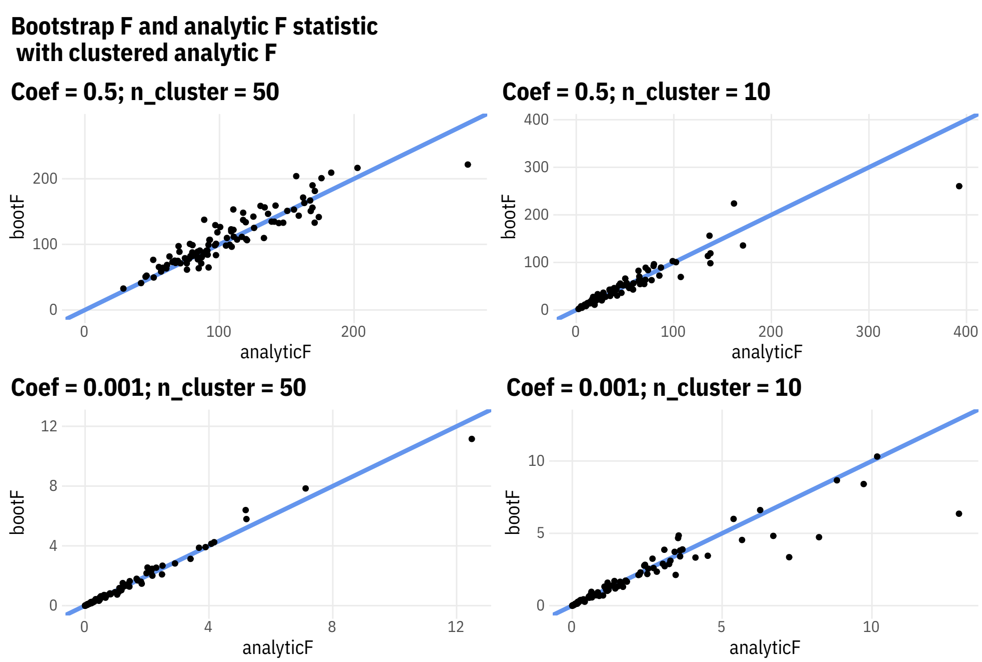
As seen in panel A, when robust analytic standard errors ignore the clustered structure, they vastly over-estimate the strength of the instrument relative to the block-bootstrap, with both "few" (10) and "many" (50) clusters and with "strong" ($\pi = 0.5$) and "weak" ($\pi = 0.001$) instruments. With appropriate clustered analytic SEs, however, the F statistic is typically comparable to the bootstrap based equivalent (panel B), although the bootstrap F is marginally more conservative with a small number of clusters and weak instrument.

In summary, we find that cluster-bootstrap F statistic and the cluster-robust F statistic, which is equivalent to the "effective" F (Olea and Pflueger, 2013) in just-identified settings such as this one, are comparable in detecting weak instruments, and recommend reporting these statistics in applied settings. We also recommend reporting Anderson-Rubin confidence intervals for the IV coefficient, as it is robust to arbitrarily weak instruments (Andrews, Stock and Sun, 2019; Kang et al., 2020).

Figure A5. Comparisons of F Statistics



(a) Cluster-bootstrap F statistic vs. Huber-White (non-clustered) F statistic



(b) Cluster-bootstrap F statistic vs. cluster-robust analytic F statistic ($F_{\texttt{Eff}}$)

A.2.2. Explaining the 2SLS-OLS Discrepancy

In this section, we conduct Monte Carlo exercises to explore potential causes of the discrepancy between 2SLS and OLS estimates observed in the replication data. We consider three causes: (1) violations of the exclusion restriction (A2), (2) publication bias, and (3) heterogeneous treatment effects (HTE). Below is our data-generating process (DGP):

$$
\begin{aligned}
& y_i = 5 + \beta_i x_i + \mu z_i + u_i + b_i \\
& x_i^* = \delta_i z_i + (1-\delta_i) a_i + 0.2 v_i \quad \text{and} \quad \delta_i = \max(\min(\kappa_i \pi_i, 1), 0) \\
& x_i = x_i^*,\ z_i \overset{i.i.d.}{\sim} N(0,2) \qquad \text{(continuous-continuous case)} \\
\text{or} \quad & x_i = 1\{x_i^* > 0\},\ z_i \overset{i.i.d.}{\sim} \text{Bern}(0.5) \qquad \text{(binary-binary case)}
\end{aligned}
$$

in which z is the instrument, x is the treatment, and y is the outcome. We consider two scenarios: (1) both x and z are continuous, and (2) both are binary. Correlated errors $\begin{bmatrix} u_i \\ v_i \end{bmatrix} \overset{i.i.d.}{\sim} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right)$; $a_i \overset{i.i.d.}{\sim} N(0,1)$, $b_i \overset{i.i.d.}{\sim} N(0,1)$ are i.i.d. errors. We use κ to control the strength of the instrument. HTE can be generated by $\begin{bmatrix} \beta_i \\ \pi_i \end{bmatrix} \overset{i.i.d.}{\sim} N\left(\begin{bmatrix} 2 \\ 1 \end{bmatrix}, \sigma_h^2 \begin{bmatrix} 1 & \lambda \\ \lambda & 0.5 \end{bmatrix}\right)$, in which σ_h controls the amount of heterogeneity in β_i and π_i while λ controls the correlation between the first stage and reduced form coefficients. δ_i is limited to be in $[0,1]$. When $\lambda > 0$, it means that a unit's treatment effect is positively correlated with its responsiveness to the IV.[A1] The sample size is fixed at 200.

Under constant treatment effect ($\sigma_h = 0$) and with a valid instrument ($\mu = 0$), the expected value of $\hat{\beta}_{2SLS}/\hat{\beta}_{OLS}$ is 0.74 for the continuous-continuous case and 0.57 for the binary-binary case. We consider four scenarios sequentially:

1. Violations of Assumption 2 are captured by $\mu \neq 0$ (failure of the exclusion restriction).

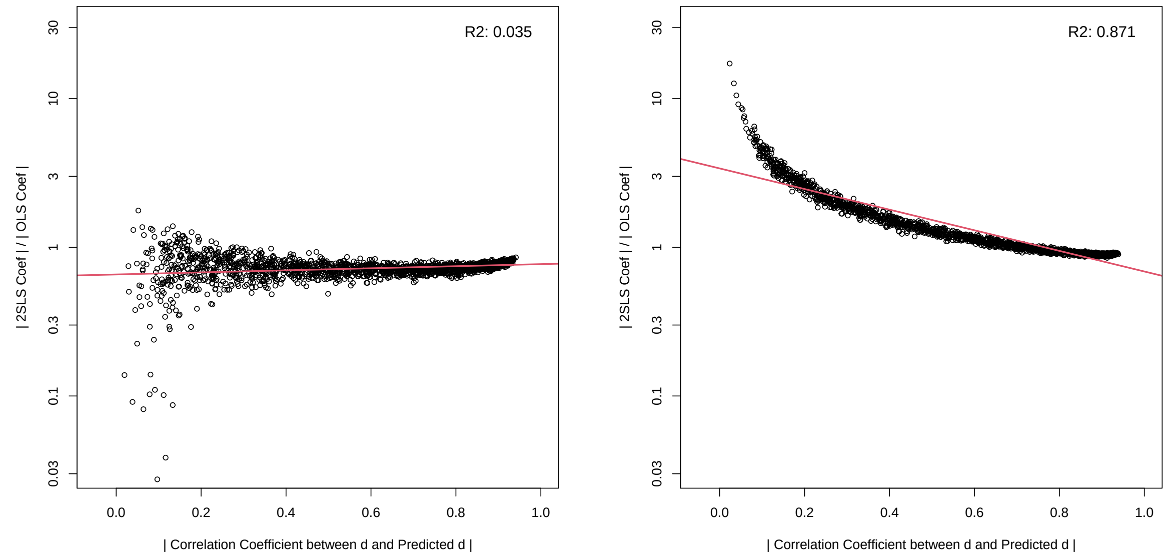
2. Publication bias can be simulated by dropping the cases in which the 2SLS estimates are statistically insignificant at the 5% using a conventional t test.

3. HTE is generated by setting $\sigma_h = 0.05$ and $\lambda = 0.7$, i.e., β_i and π_i are highly correlated.

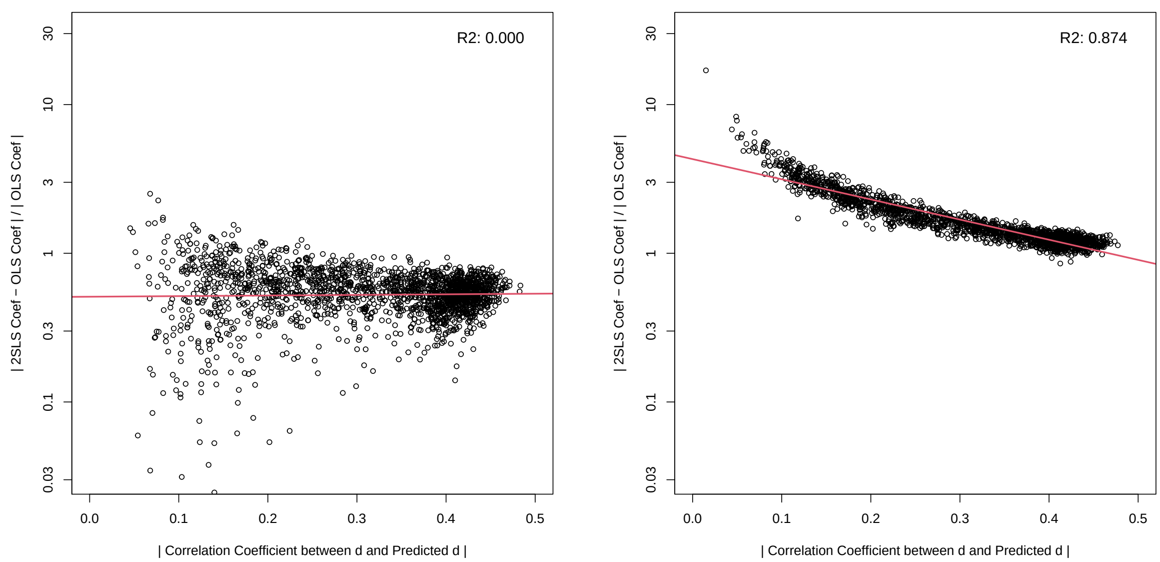
4. The combination of HTE and publication bias.

[A1]For example, under selection-on-gains type settings, which are typically considered in generalized Roy models underlying MTE approaches to IV.

Figure A6. Consequences of Exclusion Restriction Failure under Constant Effect



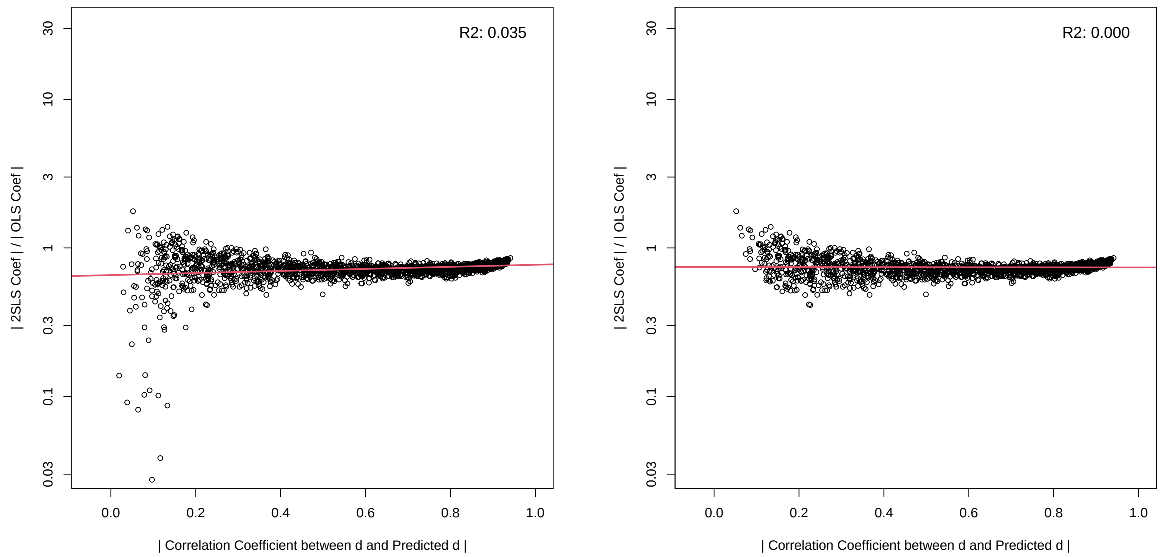
(a) Continuous-Continuous Case: w/o and w/ exclusion restriction failure



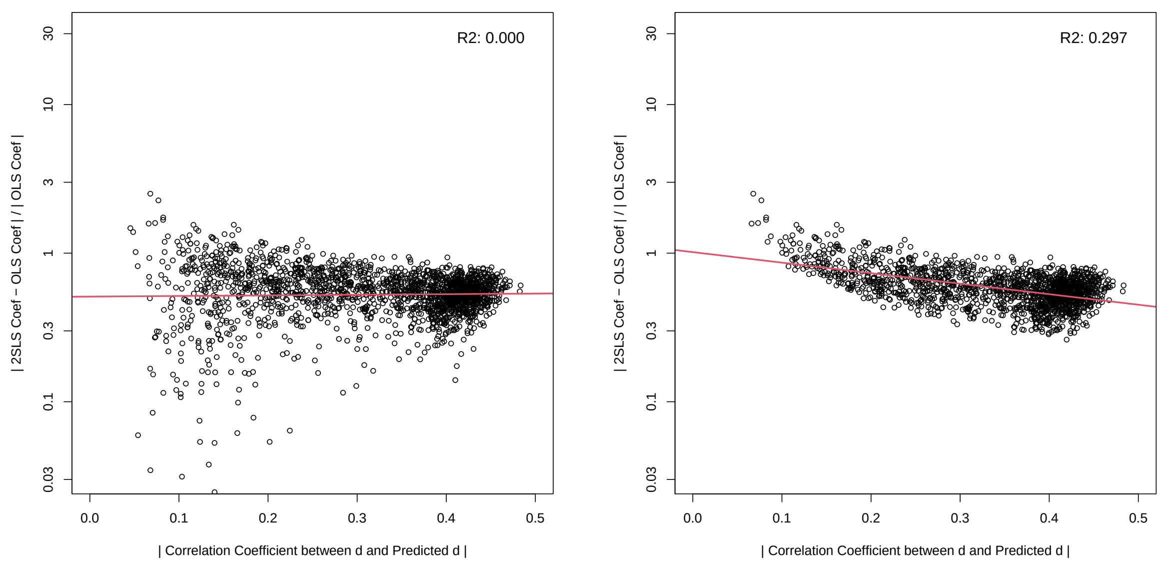
(b) Binary-Binary Case: w/o and w/ exclusion restriction failure

Violating Assumption 2. The results for Scenario 1 are shown in Figure A6. Each dot represents one simulated sample. Figure A6 shows that, in both continuous-continuous and binary-binary setups, when the treatment effect is constant ($\beta_i = \beta, \pi_i = \pi$), in expectation, there is no mechanical negative relationship between the correlation coefficient between d and $\hat{d}$ and the 2SLS-OLS discrepancy (left panels in both subfigures). However, when the exclusion restriction fails, e.g., $\mu = 1$ right panels in both subfigures), a strong negative correlation appears. These results support our argument in the paper that a weak first stage amplifies the bias from the failure of Assumption 2.

Figure A7. Consequences of Publication Bias under Constant Treatment Effect



(a) Continuous-Continuous Case: w/o and w/ publication bias

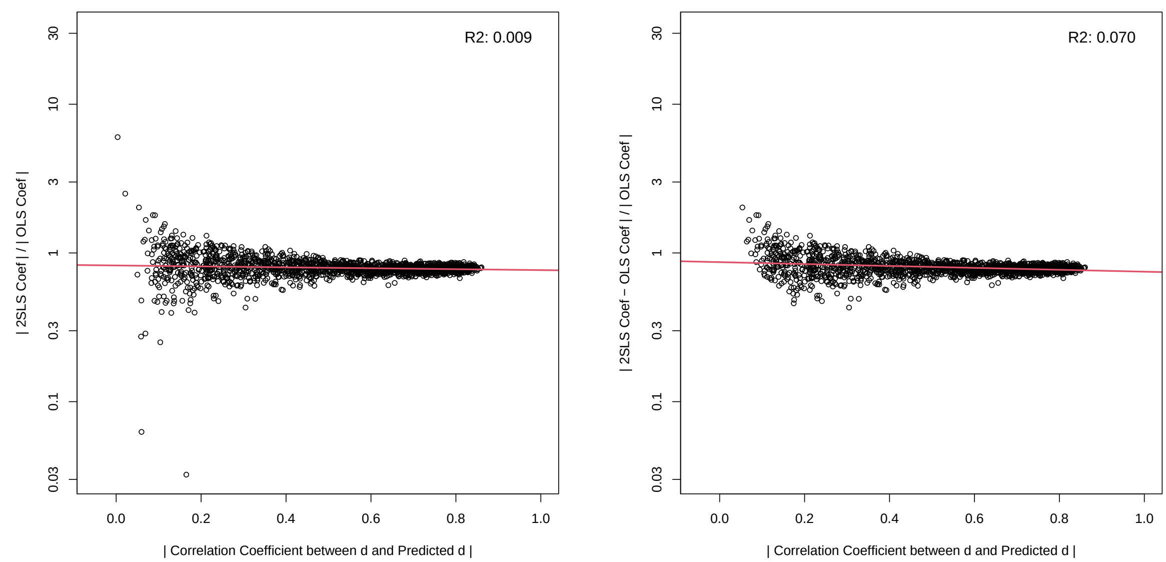


(b) Binary-Binary Case: w/o and w/ publication bias

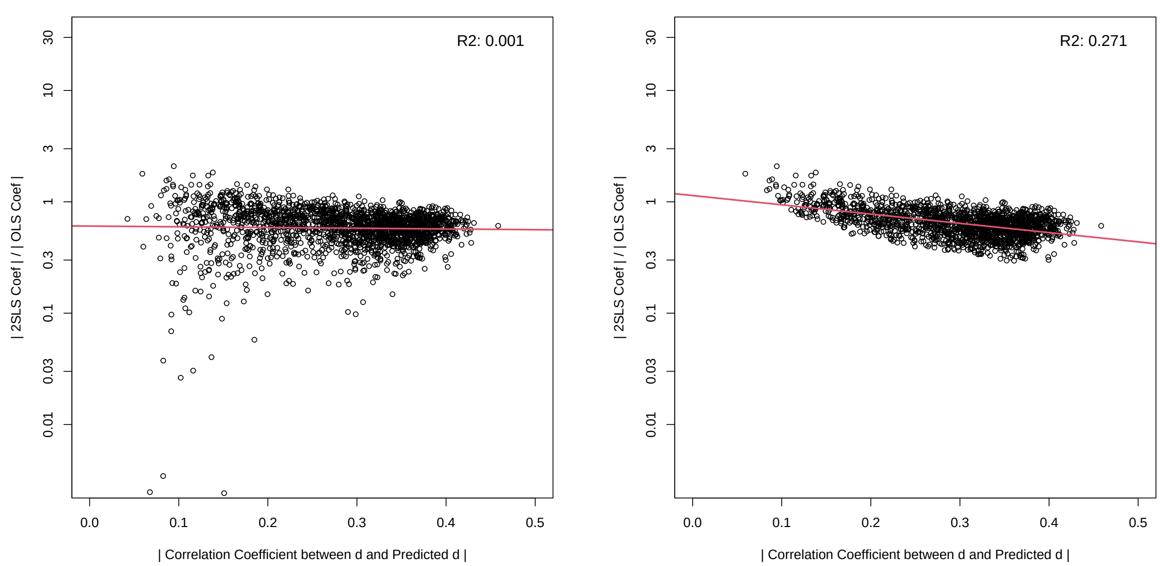
Publication bias. Figure A7 illustrates the consequences of publication bias (Scenario 2), where statistically insignificant results are omitted (right panels). The left panels are identical to the left panels in Figure A6. In the binary-binary case, we observe a moderate negative correlation; however, this correlation is much weaker than those caused by exclusion restriction failures.

HTE and publication bias. Finally, we investigate the consequences of HTE (Scenario 3) and its interaction with publication bias (Scenario 4). Figures A8 shows results under HTE, i.e., $\sigma_h > 0$ and $\lambda = 0.7$ (β_i and π_i are highly positively correlated). On the logarithmic scale, the correlation is almost nonexistent (left panels in Figure A8). When we

FIGURE A8. CONSEQUENCES OF PUBLICATION BIAS UNDER HTE



(a) Continuous-Continuous Case: w/o and w/ publication bias



(b) Binary-Binary Case: w/o and w/ publication bias revert to the original scale, we do observe a small to moderate negative correlation in both continuous-continuous and binary-binary cases (figures not shown). When we further introduce publication bias, we begin to see weak negative correlations between the first stage ρ and the 2SLS-OLS discrepancy on the logarithmic scale, especially in the binary-binary case. However, their magnitudes are much smaller than what we observed in Figure A6 under the exclusion restriction failure. This suggests that the observed strong negative relationship in the paper is unlikely to be solely explained by HTE and different levels of responsiveness to the IV.

In summary, the Monte Carlo exercises demonstrated that the strong negative correlations between the first stage ρ and the 2SLS-OLS discrepancy are most likely caused by violations of Assumption 2. Other factors, such as publication bias and HTE, may also play a role.

A.3. Evaluating the Exogeneity Assumption

Assumption 2 is a strong and generally untestable assumption that underlies the validity of the instrument; indeed, researchers typically spend considerable effort arguing for both unconfoundedness and the exclusion restrictions in their particular setting. However, some placebo tests have recently become popular as a way to argue for the validity of identification assumptions in causal designs (Eggers, Tuñón and Dafoe, 2021), especially in observational settings where the choice of IV is guided by detailed domain knowledge. Bound and Jaeger (2000) suggest first using an auxiliary regression on a subsample where the IV is not expected to influence treatment assignment, known as "zero-first-stage" (ZFS) tests. The primary intuition is that in a subsample that one has a strong prior that the first stage is zero—hence, they are "never takers," to use the language of the LATE framework—the reduced form effect should also be zero if Assumption 2 is satisfied. In other words, motivated by a substantive prior that the first-stage effect of the IV is likely zero for a subsample of the population (henceforth, the "ZFS subsample"), the researcher then proceeds to show that the reduced-form coefficient for the IV (by regression Y on Z) is approximately zero *in the ZFS subsample*, which is suggestive evidence in favor of IV validity. Most observational instruments ought to yield some ZFS subsample based on substantive knowledge of the assignment mechanism.

This style of placebo is particularly popular in studies of historical political economy, where particular historical or geographic features are argued to be valid instruments for treatment assignment, and thus they are unlikely to be driving treatment assignment outside of a specific context. For example, Nunn (2008) studies the effects of the slave trade on modern-day development in Africa using sailing distance from each country to the nearest locations of demand for slave labor as an IV for the normalized number of slaves taken. The author then argues that distance to demand locations in the New World are likely to be a valid IV by using a placebo test that the first-stage effect (the IV regressed on the outcome, modern-day GDP) is approximately zero for countries outside Africa, where the posited mechanism (that places close to demand locations exported more slaves only in the transatlantic slave trade) has no traction, thereby providing a candidate ZFS sample. In a related paper, Nunn and Wantchekon (2011) use the same strategy to show that distance to slave-trade ports does not predict modern-day trust attitudes in the Asiabarometer, while they do in the Afrobarometer (which is the primary study population). Acharya, Blackwell and Sen (2016) perform a similar exercise where they believe that their instrument (cotton suitability) predicts the treatment (slaves per capita) in the Southern States but not the

Northern states, and therefore find that the reduced form effect of cotton suitability on modern day racial attitudes is approximately zero in the Northern states.

A.3.1. The ZFS Test and Modified Inference

While this is a useful heuristic check that we advise most observational IV papers adopt, it is an informal test and provides no debiasing procedure to correct potentially biased IV estimates. Van Kippersluis and Rietveld (2018) suggest that the ZFS test can be fruitfully combined with the "plausibly exogenous" method suggested by Conley, Hansen and Rossi (2012) (henceforth, CHR 2012). To illustrate the method, we first rewrite the IV simultaneous equations in CHR (2012)'s notation:

$$Y = X\beta + Z\gamma + \varepsilon; \quad X = Z\Pi + \nu, \tag{A1}$$

where Z also enters the structural equation, and the exclusion restriction amounts to a dogmatic prior that $\gamma = 0$. CHR (2012) suggest that this assumption can be relaxed, and replaced with a user-specified assumption on a plausible value, range, or distribution for γ depending on the researcher's beliefs regarding the degree of exclusion restriction violation. They propose three different approaches for inference that involve specifying the range of values for γ, a prior distributional assumption for γ, and a fully Bayesian analysis that requires priors over all model parameters and corresponding parametric distributions. We focus on the second method, which CHR (2012) call the "local to zero" (LTZ) approximation because of its simplicity and transparency. The LTZ approximation considers "local" violations of the exclusion restriction[A2] and requires a prior over γ alone. CHR (2012) show that replacing the standard assumption that $\gamma = 0$ with the weaker assumption that $\gamma \sim \mathbb{F}$, a prior distribution, implies distribution for $\widehat{\beta}$ in Equation (A2).

$$\widehat{\boldsymbol{\beta}} \sim^a \mathcal{N}\left(\boldsymbol{\beta}, \mathbb{V}_{2SLS}\right) + \mathbf{A}\gamma \;\text{ where } \mathbf{A} \equiv (\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z} \tag{A2}$$

$$\widehat{\boldsymbol{\beta}} \sim^a \mathcal{N}\left(\boldsymbol{\beta} + \mathbf{A}\mu_\gamma, \mathbb{V}_{2SLS} + \mathbf{A}\boldsymbol{\Omega}\mathbf{A}'\right) \tag{A3}$$

where the original 2SLS asymptotic distribution is inflated by the additional term. While a simulation-based approach can be used to implement Equation (A2) for an arbitrary distribution for γ, the distribution takes its most convenient form when one uses a Gaussian prior over $\gamma \sim \mathcal{N}\left(\mu_\gamma, \Omega_\gamma\right)$, which simplifies Equation (A2) to Equation (A3), with a posterior being a Gaussian centered at $\boldsymbol{\beta} + \mathbf{A}\mu_\gamma$.

[A2] LTZ asymptotics consider a sequence of constants $\gamma = C/\sqrt{N}$ for some constant C and sample size N

CHR (2012) suggest that researchers use domain knowledge to choose $\mu_\gamma, \Omega_\gamma$, since they often hold strong priors about instruments anyway (which presumably motivates the choice of the instrument). Van Kippersluis and Rietveld (2018) suggest that a principled method to choose μ_γ is to estimate Equation (A1) on the ZFS population (wherein Π is assumed to be zero), and use this estimate $\widehat{\gamma}_{ZFS}$ as μ_γ. This approach combines the informal ZFS test with the plausibly exogenous method in a straightforward manner, and software to implement it is available in both `R` (accompanying this paper) and `STATA` (Clarke, 2014). We begin with a simulation-based illustration and illustrate the application of this method to a published empirical paper next.

A.3.2. Simulation Evidence

In this subsection, we demonstrate the LTZ method when the exclusion restriction is not satisfied. Consider the following DGP,

$$
\begin{aligned}
Y_i &= \beta_i D_i + \gamma Z_i + \varepsilon_i \\
D_i &= \mathbf{1}\{D_i^* > 0\} \\
D_i^* &= \alpha_i + \pi_i Z_i + \varepsilon_i
\end{aligned}
$$

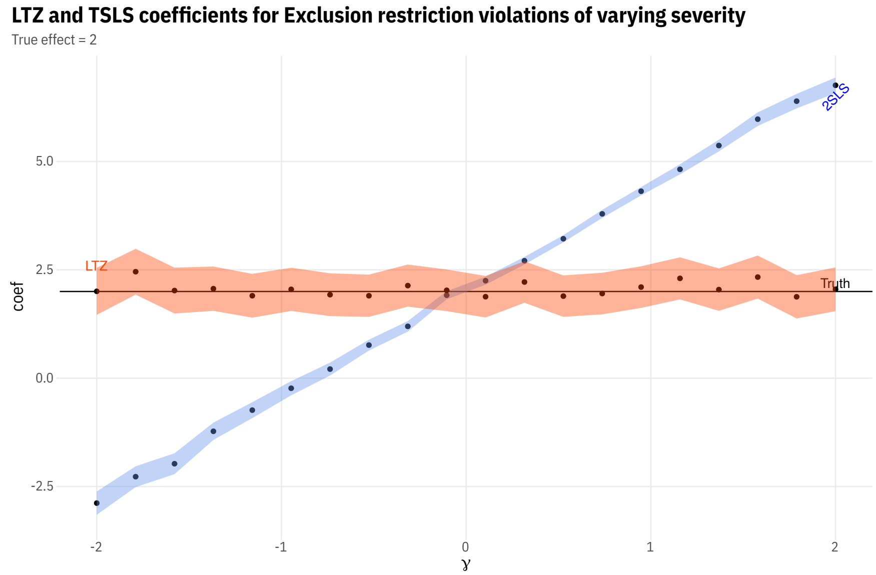
in which $Z_i \sim$ Bernoulli(0.5) is a binary instrument, $\pi_i \sim \mathsf{U}[1.5, 2.5]$, $\alpha_i \sim \mathcal{N}(-1, 1)$, $\varepsilon_i \sim \mathcal{N}(0, 1)$, $\beta_i \sim \mathcal{N}(1, 0.25)$. We generate Y_i with Z_i directly entering the structural equation, which allows us to vary the magnitude of the exclusion restriction violation. We then estimate $\widehat{\beta}_{2SLS}$ using conventional two-staged-least-squares on this data. As we vary γ, $\hat{\beta}_{2SLS}$ is inconsistent for all values except when $\gamma = 0$. We set $\pi = 0$ for the last 20% observations of the simulated data (the ZFS subsample). We then estimate the reduced-form regression on this (known) subsample and use the coefficient as a prior for μ_γ, and compute the LTZ IV estimate.

Figure A9 shows, unlike the 2SLS estimator (blue), the LTZ estimator (orange) uncovers the true value of $\beta = 2$ even for large degrees of exclusion restriction violations (large $|\gamma|$).

A.3.3. A Case Study

We illustrate the diagnostics described above by applying it to the IV analysis in Guiso, Sapienza and Zingales (2016) (henceforth GSZ 2016), who revisit Leonardi, Nanetti and Putnam (2001)'s conjecture that Italian cities that achieved self-government in the Middle Ages have higher modern-day levels of social capital. More specifically, they study the effects

Figure A9. IV and LTZ Estimates for Varying γ



of free city-state status on social capital as measured by the number of non-profits and organ donations per capita, and a measure of whether students cheat in mathematics.

Table A1. Replication of GSZ (2016) Table 6
Reduced Form Regressions

Outcome Variables	North: Nonprofit (1)	North: Organ Donation (2)	South (ZFS): Nonprofit (3)	South (ZFS): Organ Donation (4)
Bishop (IV)	1.612	0.472	0.178	0.189
	(0.219)	(0.047)	(0.137)	(0.065)
Observations	5,357	5,535	2,175	2,178

Note: Bootstrapped SEs are in the parentheses. See Figure A4 in the SM for the original table.

GSZ (2016) use a dummy for whether the city was the seat of a bishop in the Middle Ages, based on historical accounts of coordination preceding commune formation in the Middle Ages as an IV for the "free-city experience" (Section 5). They argue that conditional on a host of geographic covariates, this IV, a bishop seat, influences contemporary social capital solely through its increasing the likelihood of commune formation. As suggestive evidence for the validity of their instrument, they estimate the reduced-form effect of medieval bishop presence of contemporary social capital measures separately in the north (where the IV is conjectured to have an effect) and the south (where it is conjectured to be irrelevant). They fail to reject the null of no effects in the south, conclude that the IV appears to have face

FIGURE A10. TABLE 6 IN GUISO, SAPIENZA AND ZINGALES (2016)

TABLE 6. Validating the instrument.

A. Regressions of civic capital in the Center–North and in the South

	Center–North sample			South sample		
	(I) Nonprofit org.	(II) Organ donation org.	(III) Cheating in mathematics	(IV) Nonprofit org.	(V) Organ donation org.	(VI) Cheating in mathematics
Ease of coordination	1.61**	0.47***	−0.66***	0.18	0.19***	−0.04
	(0.219)	(0.047)	(0.118)	(0.137)	(0.065)	(0.309)
Elevation	1.93***	−0.25***	0.94**	1.43***	−0.04	0.72
	(0.475)	(0.062)	(0.441)	(0.257)	(0.083)	(0.541)
Max difference in elevation	1.35***	0.01	0.26*	−0.08	−0.05*	0.06
	(0.219)	(0.026)	(0.143)	(0.084)	(0.029)	(0.145)
City is on the coast	−0.27	−0.08*	0.03	0.23**	−0.02	0.13
	(0.264)	(0.046)	(0.119)	(0.115)	(0.044)	(0.108)
City more than 5 km from the coast	1.10*	0.07	−0.21	0.02	−0.03	1.46
	(0.634)	(0.072)	(0.227)	(0.143)	(0.048)	(1.098)
Current population	−3.38***	1.48***	−1.85***	−9.11***	1.10*	−3.50
	(1.886)	(0.290)	(0.523)	(2.242)	(0.582)	(2.849)
Current population squared	1.03	−1.12***	1.87***	6.23***	−0.86*	4.47
	(1.423)	(0.218)	(0.480)	(1.924)	(0.469)	(2.816)
Gini income inequality index	0.08	0.04	0.04	3.49**	2.05***	−21.66***
	(0.449)	(0.076)	(0.438)	(1.505)	(0.547)	(5.646)
Gini inequality index of land ownership	9.83***	2.17***	−8.61***	1.61***	0.35***	1.75
	(1.883)	(0.377)	(2.382)	(0.351)	(0.098)	(1.330)
Observations	5,357	5,535	1,911	2,175	2,178	1,210
R^2	0.083	0.587	0.023	0.329	0.574	0.027

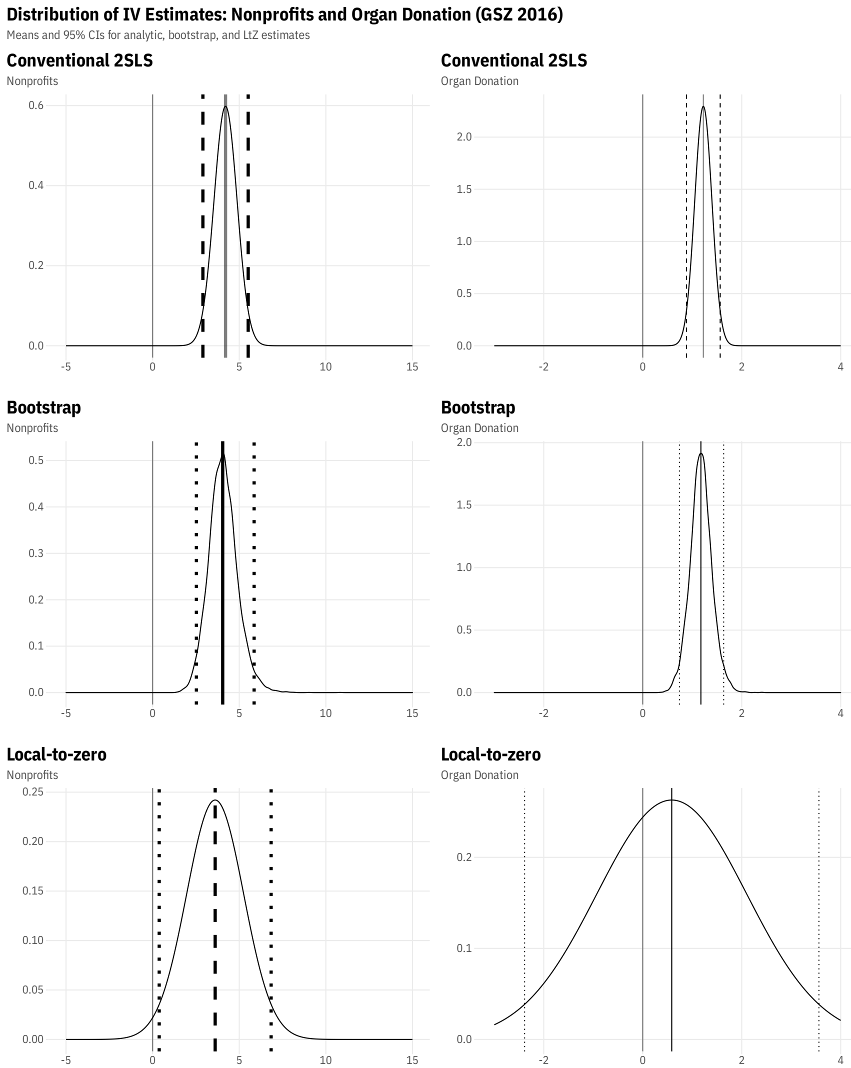
Note: "Ease of coordination" is the IV "Bishop in city." We replicated columns (I), (II), (IV), and (V).

validity, and proceed to use bishop presence as an IV for their IV estimates.

We begin by calculating the first-stage partial F statistic based on bootstrapped SEs for the north sample, which is 67.3. Because there were no "free cities" in the south, the F statistic for the south is zero by definition. We then replicate their reduced-form estimates in Table A1. The separate north and south reduced-form estimates in GSZ (2016) can be readily used for the LTZ test described above. The authors substantively believe that the south is a ZFS sample where bishop presence is irrelevant for treatment assignment,[A3] we can use the reduced-form estimates of 0.178 and 0.189 in the south for nonprofits per capita and organ donation (columns 3-4 in Table A1) as the prior μ_γ for the direct effect of the IV on the outcome. Finally, we report the analytic, bootstrap, and LTZ IV results in Figure A11. We

[A3]The authors claim this indirectly by reporting the reduced form effects separately for the north and south subsamples in Table 6, and state that since the reduced form is attenuated in the south, this justifies the use of bishop presence as an IV (p. 1427).

FIGURE A11. IV COEFFICIENTS FOR NON-PROFITS AND ORGAN DONATION



find that conventional robust SEs understate the uncertainty of the estimates relative to the bootstrap and that accounting for direct effect using LTZ attenuates GSZ (2016)'s estimates somewhat and substantially increases the SE of the estimate for the nonprofit outcome. For organ donation, however, where we suspect a violation of Assumption 2 because the reduced form effect is statistically distinguishable from zero, the use of the LTZ method to account for this exclusion restriction violation yields a smaller and substantially more uncertain estimate whose CI contains 0. This example shows how researchers may take advantage of the ZFS test and the LTZ technique to gauge the robustness of their findings based on an IV strategy.

A.4. Summary of Replicated Papers

Table A2. Summary of Replicated Papers

Paper	Instrument	Treatment	Outcome	IV Type	Justification for IV Validity
			APSR		
Gerber, Huber and Washington (2010)	Being sent mail	Aligning party identification with latent partisanship	Voting and party alignment scale	Experiment	NA
Meredith (2013)	Governors home county	Democratic governor	Down-ballot Democratic candidates vote share	Theory (Other)	"The validity of the instruments hinges on the assumption that, conditional on the control variables, coattail effects are the only channel through which the place of birth or residence of a partys gubernatorial candidate affects the vote shares received by its down-ballot candidates."(p.745)
Blattman, Hartman and Blair (2014)	Assignment to treatment blocks	Mass education campaign for dispute resolution	Serious land dispute	Experiment	NA
Laitin and Ramachandran (2016)	Geographic distance from the origins of writing	Language choice	Human development index	Theory (Geography)	"[T]he distance from these sites of invention should have no independent impact on socioeconomic development today, except through the channel of affecting the probability of possessing a writing tradition." (p. 470)
Ritter and Conrad (2016)	Rainfall	Mobilized dissent	Repression	Theory (Weather)	"[R]ainfall is an exogenous predictor of dissent onset, meeting the key criteria for the instrumental analysis to allow for causal inference."(p.89)
Croke et al. (2016)	Access to the secondary education	Education attainment	Political participation	Rules & policy changes (Change in exposure)	"There are, however, good reasons to believe that the secondary education reform only affects participation through its effect on educational attainment."(p.592)
Dower et al. (2018)	Level of serfdom	Frequency of unrest	Peasant representation and unrest	Theory (History)	"After conditioning on these covariates, we are left with that portion of serfdom largely determined by idiosyncratic variation in land grants to the nobility decades or centuries before the zemstvo reform of 1864." (p. 133)

Dower et al. (2018)	Religious polarization	Frequency of unrest	Peasant representation and unrest	Theory (History)	"After conditioning on these covariates, we are left with that portion of serfdom largely determined by idiosyncratic variation in land grants to the nobility decades or centuries before the zemstvo reform of 1864." (p. 133)
Nellis and Siddiqui (2018)	Narrow victory by secular parties in a district	The proportion of MNA seats in a district won by secularist candidates	Religious violence	Theory (Election)	"Our identifying assumption is that the outcomes of such close elections are as good as randomly decided." (p. 50)
Kapoor and Magesan (2018)	Changes in entry costs.	Number of independent candidates	Voter turnout	Rules & policy changes (Change in exposure)	"It is worth reiterating that the deposit increases had nothing to do with historical differences in voter and candidate participation across reserved and open constituencies." (p. 681)
Colantone and Stanig (2018*a*)	Imports from China to the United States × local industrial structure	Regional-level import shock from China	Leave support in Brexit	Econometrics (Interaction)	"[The]instrument is meant to capture the variation in Chinese imports, which is due to the exogenous changes in supply conditions in China, rather than to domestic factors in the United Kingdom that could be correlated with electoral outcomes." (p. 206)
Hager, Krakowski and Schaub (2019)	Distance to the nearest location where armored military vehicles were stolen	Ethnic riots (destruction)	Prosocial behavior	Theory (Other)	"[W]e present a falsification test which corroborates that the instrument is unrelated to prosocial behavior in a sample of 136 nearby villages, thus underlining the exclusion restriction." (p. 1037)
Baccini and Weymouth (2021)	Bartik instrument	Manufacturing Layoffs	Change of Democratic Vote Share	Econometrics (Interaction)	"Since layoffs are not randomly assigned, we develop an instrumental variables strategy using shift-share methodology (Bartik 1991) derived from national layoff shocks, weighted by initial county-level employment."(p.550)
Hager and Krakowski (2022)	Number of corrupted Catholic priests	Number of secret police officers	Resistance	Theory (History)	"In the early days of the regime, the secret polices ability to servile citizens depended critically on the cooperation of the Catholic Church...Importantly, the corruptibility of priests was plausibly exogenous: priests were sent to municipalities by the Catholic Church, often when another priest had retired." (p.565)

Kuipers and Sahn (2022)	Statewide assignment mandate	Civil service reform	Descriptive representation on an unrestricted sample	Rules & policy changes (Assignment)	“First, we assume that state-level mandates are a strong instrument for city adoption; we verify the strength of the instrument in the main presentation of the results. The exclusion restriction, which is untestable, seems a reasonable assumption in our case.”(p.9)
			AJPS		
Kocher, Pepinsky and Kalyvas (2011)	Past insurgent control	Aerial bombing	Changes in local control	Theory (Other)	“Because instrumental variables require only conditional independence between instruments and the error term, we need only assume that there are no unobserved hamlet-specific variables that affected insurgent control in July, August, and December 1969, but not in September of that year as well.” (p. 212)
Vernby (2013)	Immigration Inflow 19401950; immigration Inflow 19601967	Share of noncitizens in the electorate	Municipal education and social spending	Theory (History)	“Furthermore, it is unlikely that the initial locations of these refugees were affected by the level of local public services, suggesting that the instrument is also valid.” (p. 25)
Tajima (2013)	Distance to health station	Distance to police posts (as a proxy for exposure to military intervention)	Incidence of communal violence	Theory (Geography)	“According to a Health Department official, primary health stations must be located in every subdistrict at their population centers, regardless of the propensity for violence of those locations” (p. 112)
De La O (2013)	Random assignment to early coverage	Early coverage of Conditional Cash Transfer	Incumbent party’s vote share	Experiment	NA
McClendon (2014)	Assignment to treatment	Reading social esteem promising email	Participation in LGBTQ events	Experiment	NA
Barth, Finseraas and Moene (2015)	Adjusted bargaining coverage and effective number of union confederations	Wage inequality	Welfare support	Theory (Other)	“Yet conditional on union density and country fixed effects, we argue that certain properties of the bargaining system are likely to affect wages, but not union involvement in politics.” (p. 574)

Stokes (2016)	Wind speed	Turbine location	Vote turnout	Theory (Climate)	"Wind speed is theoretically orthogonal to precinct boundaries but predicts the placement of wind turbine locations." (p. 965)
Coppock and Green (2016)	Mailing showing 2005 Vote	Voting in November 2007 municipal elections	Voting in the 2008 presidential primary	Experiment	NA
Trounstine (2016)	The number of waterways in a city combined with logged population	Racial segregation	Direct general expenditures	Theory (Geography)	"I focus on waterways (including large streams and rivers), which vary in number across cities and are arguably exogenous to segregation and spending." (p. 717)
Carnegie and Marinov (2017)	Being a former colony of one of the Council members	Foreign aid	CIRI Human Empower-ment index	Theory (History)	"In 1965, the EU stipulated that countries would hold the presidency for 6 months at a time [...] and would rotate alphabetically according to each member states name as spelled in its own language. " (p. 676)
Zhu (2017)	Weighted geographic distance from economic centers	MNC activity	Corruption	Theory (Geography)	"This instrumental variable (IV) is rooted in the gravity models of international trade and FDI flows." (p. 90)
Rueda (2017)	The size of the polling station	Actual polling place size	Citizens' reports of electoral manipulation	Rules & policy changes (Fuzzy RD)	"The institutional rule predicts sharp reductions in the size of the average polling station of a municipality every time the number of registered voters reaches a multiple of the maximum number of voters allowed to vote in a polling station." (p. 173)
Lelkes, Sood and Iyengar (2017)	State-level ROW index	Number of providers	Affective polarization (partisan hostility)	Theory (Other)	"[A]n index of state regulation of right-of-way laws strongly predicts the number of providers in a county, which, as we discuss later, is a good proxy for broadband uptake." (p. 4).
Goldstein and You (2017)	Direct flight from city to Washington DC	Lobbying spending	Total earmarks or grants awarded	Theory (Other)	"The existence of a direct flight captures the convenience of travel to Washington, DC, from each city." (p. 865)

Spenkuch and Tillmann (2018)	Individual princes decisions concerning whether to adopt Protestantism	Religion of voters living in the same areas more than three and a half centuries later	Nazi vote share	Theory (History)	“The historical record, however, suggests that princes decisions may plausibly satisfy this exogeneity assumption, especially after controlling for economic conditions at the end of the Weimar Republic as well as all factors known to have influenced rulers.” (p. 27)
Colantone and Stanig (2018*b*)	Chinese imports to the United States ×regional industrial structure	Regional import shock from China	Economic nationalism	Econometrics (Interaction)	“This instrument is meant to capture the variation in Chinese imports due to exogenous changes in supply conditions in China, rather than to domestic factors that could be correlated with electoral outcomes.” (p. 6)
Hager and Hilbig (2019)	Mean elevation	Equitable inheritance customs	Female representation	Theory (Geography; History)	“Rivers are exogenous, but no longer should have a strong effect on inequality other than through the treatment.” (p. 767)
Hager and Hilbig (2019)	Distance to rivers	Equitable inheritance customs	Female representation	Theory (Geography; History)	“Rivers are exogenous, but no longer should have a strong effect on inequality other than through the treatment.” (p. 767)
Chong et al. (2019)	Treatment assignment in get-out-to-vote campaigns	Actual proportion of households treated in the locality	Voted in 2013 presidential election	Experiment	NA
Kim (2019)	Population threshold	Democratic institutions	Women political engagement	Rules & policy changes (Fuzzy RD)	“[L]ocalities with a population greater than 1,500 must create a municipal council [...] whereas those with a population below that threshold were free to choose between the status quo direct democracy and representative democracy.” (p. 6).
Sexton, Wellhausen and Findley (2019)	Soldier fatalities	Health budget	Welfare outcome	Theory (Other)	“We substantiate [the exclusion restriction] below by ruling out the key alternative channel that local insecurity could affect citizens use of health services.” (p. 359)
López-Moctezuma et al. (2022)	Assignment to treatment	Town-hall meetings	Voting behavior	Experiment	NA

Blair, Di Salvatore and Smidt (2022)	Average fragmentation of all ongoing PKO mandates	Fragmentation of any given PKO mandate	Process performance	Theory (Other)	"We view the first of these assumptions as mostly uncontroversial. As discussed above, most PKO mandates are only loosely tailored to conditions in their host countries. It is highly unlikely that the mandates of all other PKOs in Africa are tailored to the host country conditions of any given PKO. This should mitigate independence concerns." (p.11)
Hong, Park and Yang (2022)	Geographic terrain elevation and slope	NVM subsidies	Parks vote share in 2012	Theory (Geography)	"The logic behind this choice is as follows: each villages performance in the NVM is evaluated based on their baseline conditions. Therefore, an unfavorable terrain before the movement likely indicates an initial lack of infrastructure in a poorer environment, and thus gives a village an advantageous benchmark from which to generate a notable and visible improvement within a short period compared to other villages." (p.11)
Wood and Grose (2022)	Random audit	Incumbent found to have campaign finance violations	Legislator retired	Experiment	NA
			JOP		
Gehlbach and Keefer (2012)	Whether the first ruler in a nondemocratic episode is a military leader	Age of ruling party less leader years in office	Private investment/GDP	Theory	"[D]ictators who come to power with the backing of the military require less popular support to remain in power and are therefore less likely to promote private investment by allowing supporters to organize." (p. 628)
Healy and Malhotra (2013)	Whether the younger sibling is a sister	The share of a respondents siblings who are female	1973 gender-role attitude	Theory (Others – Biology)	"However, under Assumption 1, all siblings have an impact only through the overall gender makeup of the household." (p. 1027)
Dube and Naidu (2015)	US military aid to countries outside of Latin America	US military aid to Colombia	The number of paramilitary attacks	Theory (Diffusion)	"The instrument is valid since US funding to the rest of the world is determined by the broad geopolitical outlook of the American government, reflecting factors such as the party of the president or other major world events, and can thus be considered exogenous to the conflict in Colombia."(p.256)

Flores-Macias and Kreps (2013)	Lagged values of country's energy production	Trade volume	Foreign policy convergence	Theory (Other)	"The logic is that trade and trade salience in Africa and Latin America are significantly related to countries energy production, but there is no reason to believe that either of them is correlated with the error term in the equation predicting foreign policy convergence" (p. 365)
Charron and Lapuente (2013)	Consolidation of clientelistic networks in regions where rulers have historically less constraints to their decisions	Clientelism	Quality of government	Theory (History)	"[W]e also find that constraints are directly correlated with current regional institutional quality (yet in his analysis regional GDP and GDP growth are used), thus rendering it an imperfect instrument for clientelism"(p.576)
Kriner and Schickler (2014)	Number of days Congress is in session	Committee investigations	Presidential approval	Theory (Other)	"[T]here is no theoretical reason drawn from existing literatures to expect the calendar to be independently correlated with presidential approval." (p. 525)
Lorentzen, Landry and Yasuda (2014)	Large firm dominance in 1999	Large firm dominance in 2007	Pollution information transparency index	Econometrics (Lagged treatment)	"[The instrument was measured] well before transparency reforms were a major focus of discussion." (p. 187)
Dietrich and Wright (2015)	Constructed "internal" excluded instrument	Economic aid	Transitions to multipar-tyism	Econometrics (Lewbel instrument)	"[We]show that the excluded instruments are generally uncorrelated with alternative channels through which they might influence the outcome variables." (p. 223)
Feigenbaum and Hall (2015)	Chinese exports to other economies $\times$local exposure	Localized trade shocks in congressional districts	Trade score	Econometrics (Interaction)	"[We] use an instrument that depends [...] on Chinese import growth to other rich, Western economies" and "the lagged version is unaffected by Chinese trade shock." (p.1019)
Alt, Marshall and Lassen (2016)	Assignment to receiving an aggregate unemployment forecast	Unemploy-ment expectations	Vote intention	Experiment	NA
Johns and Pelc (2016)	Trade stake of the rest of the world	The number of other countries that became third parties	Becoming a third party	Theory (Other)	"[E]ach states participation decision is not directly affected by the trade stake of other countries. The trade stake of other countries matters only to the extent that it shapes a players belief about how other countries will behave." (p. 99)

Acharya, Blackwell and Sen (2016)	Measures of the environmental suitability for growing cotton	Slave proportion in 1860	proportion Democrat	Theory (History)	“We present results from this analysis showing that, outside the South, the relationship between cotton suitability and political attitudes is either very small or in the opposite direction as in the South.” (p. 628)
Schleiter and Tavits (2016)	Prime Minister dissolution power	Opportunis-tic election calling	Vote share of Prime Minister’s party	Theory (Other)	“The instrument correlates directly with the treatment of interestopportunistic election callingwithout being linked to anticipated incumbent electoral performance.” (p. 840)
Henderson and Brooks (2016)	Rain around Election day	Democratic vote margins	Incumbent roll call positioning	Theory (Weather)	“Rain several days before an election may dampen the willingness to make plans, arrange transportation, and schedule time off work to go to the polls.”(p.657)
Henderson and Brooks (2016)	Rain around Election weekend	Democratic vote margins	Incumbent roll call positioning	Theory (Weather)	“Rain several days before an election may dampen the willingness to make plans, arrange transportation, and schedule time off work to go to the polls.”(p.657)
Charron et al. (2017)	Proportion of Protestant residents in a region; aggregate literacy in 1880	More developed bureaucracy	Percent of single bidders	Theory (History)	“[C]ross-country data show that, while the least corrupted countries in the world all have had near universal literacy for decades, other countries considered highly corrupt, [...] have, for the entire postwar era, also been some of the most highly literate places in the world.” (p.97)
West (2017)	IEM (prediction market) price	Obama win	Policy efficacy	Theory (Other)	“The identifying assumption is that there is no unobservable factor that simultaneously affects black (female) political efficacy and perceptions of the likelihood of an Obama (Clinton) victory.” (p.352)
Stewart and Liou (2017)	Log total border length and the total number of that states neighbors	Foreign territorial control	Civilian casualties	Theory (Geography)	“[T]he longer a states borders or the greater its number of neighbors, the more accessible border regions in neighboring states will be to rebels, independent of the dynamics of their conflict with the government. Further, total border length or the number of bordering states is not likely to affect rebel targeting of civilians other than through their effects on the likelihood of rebel groups controlling foreign territory.” (p. 291)
Lerman, Sadin and Trachtman (2017)	Born 1946 or 1947	Public (p 1) versus only private (p 0) health insurance	Support ACA	Rules & policy changes (Change in exposure)	“We can confirm across a host of observable covariates that these two age groups are similar on almost every dimension, with the exception of insurance.” (p. 631)

Grossman, Pierskalla and Boswell Dean (2017)	The number of distinct landmasses; length of medium and small streams; over-time variation in the number of regional governments	Government fragmentation	Public goods provision	Theory (Geography / diffusion)	"Territorial structure of neighboring countries will affect the local discourse on institutional reforms and increase the likelihood that a country will adopt similar reforms" and "The other two instruments build on the fact that administrative and political boundaries are drawn around geographic landmarks." (p. 831)
Cirone and Van Coppenolle (2018)	Random assignment of budget incumbents to bureaux	Budget committee service	Legislator sponsorship on a budget bill	Theory (Other)	"Conceptually, the competitiveness of the randomly assigned group acts similarly to a form of encouragement design." (p. 953)
Bhavnani and Lee (2018)	Early-career job assignment to districts	Bureaucrats embeddedness	Proportion of villages with high schools	Theory (Other)	"[T]he IAS posting orders that we obtained suggest that heuristics such alphabetical order and serial numberwhich are arbitrary and orthogonal to district and officer characteristicsare used to match officers to districts." (p. 78)
Pianzola et al. (2019)	Random assignment of the e-mail treatment	Smartvote use	Vote intention	Experiment	NA
Arias and Stasavage (2019)	Trade shock × UK bond yield	Government expenditures	Regular leader turnover	Econometrics (Interaction)	"The logic here is that when costs of external borrowing are high, a government experiencing a trade shock is more likely to cut expenditures because the option of borrowing to maintain or increase expenditures is too costly. This interaction term is the excluded instrument while the Trade Shock variable is included in both the first- and the second-stage estimates" (p. 1519)
Ziaja (2020)	Constructed instrument	Number of democracy donors	Democracy scores	Econometrics (Interaction)	"[T]here is no reason to believe that the gender composition of a donor countrys parliament should affect democracy in a recipient country directly."(p.439)
Schubiger (2021)	counterinsurgent mobilization	exposure to state violence	Location of a community inside or outside the emergency zone	Theory(Geography)	"Destination choices were typically driven by economic and social factors (e.g., Degregori 1998, 151; Del Pino 1996, 164). Moreover, it is unlikely that local residents were able to anticipate the boundaries of the emergency zones and whether, when, and where they would change over time." (p.1389)

DiGiuseppe and Shea (2022)	Echelon corridor	US support	Fiscal capacity	The-ory(Geography)	“Like Aklin and Kern (2019), we nd that the echelon is plausibly exogenous to a states capacity, property rights, or risk of conict. Instead, whether a state is located in the echelon corridor is a function of happenstance geography.”(p.777)
Lei and Zhou (2022)	Whether the city has more than 3 million residents	Subway approval	Mayor promotion	Rules & policy changes (Fuzzy RD)	“the citys population exceeds 3 million people, and (4) more than 30,000 people per hour are expected to use a subway line”(p.463)
Urpelainen and Zhang (2022)	Time trend multiplied by the wind resource of the electoral district	Wind turbine capacity	Democratic vote	Economet-rics(Interaction)	“Validity of the average wind resource instrument hinges on two criteria: relevance and exclusion restriction...”(pp.1313-1314)
Webster, Connors and Sinclair (2022)	Treatment assignment	Percentage of angry words that a respondent wrote in emotional recall prompts	Social polarization	Experiment	NA

Note: Justifications are omitted in the case of randomized controlled trials.

References

Acharya, Avidit, Matthew Blackwell and Maya Sen. 2016. "The political Legacy Of American Slavery." *The Journal of Politics* 78(3):621–641.

Alt, James E., John Marshall and David D. Lassen. 2016. "Credible Sources And Sophisticated Voters: When Does New Information Induce Economic Voting?" *The Journal of Politics* 78(2):327–342. Publisher: University of Chicago Press Chicago, IL.

Andrews, Isaiah, James Stock and Liyang Sun. 2019. "Weak Instruments In Instrumental Variables Regression: Theory And Practice." *Annual Review of Economics* 11:727–753.

Arias, Eric and David Stasavage. 2019. "How Large Are The Political Costs Of Fiscal Austerity?" *The Journal of Politics* 81(4):1517–1522.

Baccini, Leonardo and Stephen Weymouth. 2021. "Gone for good: Deindustrialization, white voter backlash, and US presidential voting." *American Political Science Review* 115(2):550–567.

Barth, Erling, Henning Finseraas and Karl O. Moene. 2015. "Political Reinforcement: How Rising Inequality Curbs Manifested Welfare Generosity." *American Journal of Political Science* 59(3):565–577. Publisher: Wiley Online Library.

Bhavnani, Rikhil R. and Alexander Lee. 2018. "Local Embeddedness And Bureaucratic Performance: Evidence From India." *The Journal of Politics* 80(1):71–87. Publisher: University of Chicago Press Chicago, IL.

Blair, Robert A, Jessica Di Salvatore and Hannah M Smidt. 2022. "When do UN peacekeeping operations implement their mandates?" *American Journal of Political Science* 66(3):664–680.

Blattman, Christopher, Alexandra C. Hartman and Robert A. Blair. 2014. "How To Promote Order And Property Rights Unde rWeak Rule Of Law? An Experiment In Changing

Dispute Resolution Behavior Through Community Education." *American Political Science Review* p. 100–120. Publisher: JSTOR.

Bound, John and David A Jaeger. 2000. "Do Compulsory School Attendance Laws Alone Explain The Association Between Quarter Of Birth And Earnings?" *Research in labor economics* 19(4):83–108.

Carnegie, Allison and Nikolay Marinov. 2017. "Foreign Aid, Human Rights, And Democracy Promotion: Evidence From A Natural Experiment." *American Journal of Political Science* 61(3):671–683. Publisher: Wiley Online Library.

Charron, Nicholas, Carl Dahlström, Mihaly Fazekas and Victor Lapuente. 2017. "Careers, Connections, And Corruption Risks: Investigating The Impact Of Bureaucratic Meritocracy On Public Procurement Processes." *The Journal of Politics* 79(1):89–104. Publisher: University of Chicago Press Chicago, IL.

Charron, Nicholas and Victor Lapuente. 2013. "Why Do Some Regions In Europe Have A Higher Quality Of Government?" *The Journal of Politics* 75(3):567–582. Publisher: Cambridge University Press New York, USA.

Chong, Alberto, Gianmarco León-Ciliotta, Vivian Roza, Martín Valdivia and Gabriela Vega. 2019. "Urbanization Patterns, Information Diffusion, And Female Voting In Rural Paraguay." *American Journal of Political Science* 63(2):323–341. Publisher: Wiley Online Library.

Cirone, Alexandra and Brenda Van Coppenolle. 2018. "Cabinets, Committees, And Careers: The Causal Effect Of Committee Service." *The Journal of Politics* 80(3):948–963. Publisher: University of Chicago Press Chicago, IL.

Clarke, Damian. 2014. "PLAUSEXOG: Stata Module To Implement Conley Et Al's Plausibly Exogenous bounds." Statistical Software Components, Boston College Department of Economics.

URL: *https://ideas.repec.org/c/boc/bocode/s457832.html*

Colantone, Italo and Piero Stanig. 2018*a*. "Global Competition And Brexit." *American political science review* 112(2):201–218.

Colantone, Italo and Piero Stanig. 2018*b*. "The Trade Origins Of Economic Nationalism: Import Competition And Voting Behavior In Western Europe." *American Journal of Political Science* 62(4):936–953. Publisher: Wiley Online Library.

Conley, Timothy G, Christian B Hansen and Peter E Rossi. 2012. "Plausibly Exogenous." *The review of economics and statistics* 94(1):260–272.

Coppock, Alexander and Donald P. Green. 2016. "Is Voting Habit Forming? New Evidence From Experiments And Regression Discontinuities." *American Journal of Political Science* 60(4):1044–1062. Publisher: Wiley Online Library.

Croke, Kevin, Guy Grossman, Horacio A. Larreguy and John Marshall. 2016. "Deliberate Disengagement: How Education Can Decrease Political Participation In Electoral Authoritarian Regimes." *American Political Science Review* 110(3):579–600. Publisher: Cambridge University Press.

De La O, Ana L. 2013. "Do Conditional Cash Transfers Affect Electoral Behavior? Evidence From A Randomized Experiment In Mexico." *American Journal of Political Science* 57(1):1–14. Publisher: Wiley Online Library.

Dietrich, Simone and Joseph Wright. 2015. "Foreign Aid Allocation Tactics And Democratic Change In Africa." *The Journal of Politics* 77(1):216–234. Publisher: University of Chicago Press Chicago, IL.

DiGiuseppe, Matthew and Patrick E Shea. 2022. "Us patronage, state capacity, and civil conflict." *The Journal of Politics* 84(2):767–782.

Dower, Paul Castaneda, Evgeny Finkel, Scott Gehlbach and Steven Nafziger. 2018. "Collective Action And Representation In Autocracies: Evidence From RussiaExtquoterights

Great Reforms." *American Political Science Review* 112(1):125–147. Publisher: Cambridge University Press.

Dube, Oeindrila and Suresh Naidu. 2015. "Bases, Bullets, And Ballots: The Effect Of Us Military Aid On Political Conflict In Colombia." *The Journal of Politics* 77(1):249–267. Publisher: University of Chicago Press Chicago, IL.

Eggers, Andrew C, Guadalupe Tuñón and Allan Dafoe. 2021. "Placebo Tests For Causal Inference." Mimeo, University of Chicago.

Feigenbaum, James J. and Andrew B. Hall. 2015. "How Legislators Respond To Localized Economic Shocks: Evidence From Chinese Import Competition." *The Journal of Politics* 77(4):1012–1030. Publisher: University of Chicago Press Chicago, IL.

Flores-Macias, Gustavo A. and Sarah E. Kreps. 2013. "The Foreign Policy Consequences Of Trade: ChinaExtquoterights Commercial Relations With Africa And Latin America,1992Extendash2006." *The Journal of Politics* 75(2):357–371. Publisher: Cambridge University Press New York, USA.

Gehlbach, Scott and Philip Keefer. 2012. "Private Investment And The Institutionalization Of Collective Action In Autocracies: Ruling Parties And Legislatures." *The Journal of Politics* 74(2):621–635. Publisher: Cambridge University Press New York, USA.

Gerber, Alan S., Gregory A. Huber and Ebonya Washington. 2010. "Party Affiliation, Partisanship, And Political Beliefs: A Field Experiment." *American Political Science Review* 104(4):720–744. Publisher: Cambridge University Press.

Goldstein, Rebecca and Hye Young You. 2017. "Cities As Lobbyists." *American Journal of Political Science* 61(4):864–876. Publisher: Wiley Online Library.

Grossman, Guy, Jan H. Pierskalla and Emma Boswell Dean. 2017. "Government Fragmentation And Public Goods Provision." *The Journal of Politics* 79(3):823–840. Publisher: University of Chicago Press Chicago, IL.

Guiso, Luigi, Paola Sapienza and Luigi Zingales. 2016. “Long-term Persistence.” *Journal of the European Economic Association* 14(6):1401–1436.

Hager, Anselm and Hanno Hilbig. 2019. “Do Inheritance Customs Affect Political And Social Inequality?” *American Journal of Political Science* 63(4):758–773. Publisher: Wiley Online Library.

Hager, Anselm and Krzysztof Krakowski. 2022. “Does state repression spark protests? evidence from secret police surveillance in communist poland.” *American Political Science Review* 116(2):564–579.

Hager, Anselm, Krzysztof Krakowski and Max Schaub. 2019. “Ethnic Riots And Prosocial Behavior: Evidence From Kyrgyzstan.” *American Political Science Review* 113(4):1029–1044.

Healy, Andrew and Neil Malhotra. 2013. “Childhood Socialization And Political Attitudes: Evidence From A Natural Experiment.” *The Journal of Politics* 75(4):1023–1037. Publisher: Cambridge University Press New York, USA.

Henderson, John and John Brooks. 2016. “Mediating The Electoral Connection: The Information Effects Of Voter Signals On Legislative Behavior.” *The Journal of Politics* 78(3):653–669.

Hong, Ji Yeon, Sunkyoung Park and Hyunjoo Yang. 2022. “In Strongman We Trust: The Political Legacy of the New Village Movement in South Korea.” *American Journal of Political Science* .

Johns, Leslie and Krzysztof J. Pelc. 2016. “Fear Of Crowds In World Trade Organization Disputes: Why DonExtquoterightt More Countries Participate?” *The Journal of Politics* 78(1):88–104. Publisher: University of Chicago Press Chicago, IL.

Kang, Hyunseung, Yang Jiang, Qingyuan Zhao and Dylan S Small. 2020. “IV Model: An R Package For Inference And Sensitivity Analysis Of Instrumental Variables Models With One Endogenous Variable.” *arXiv preprint arXiv:2002.08457* .

Kapoor, Sacha and Arvind Magesan. 2018. "Independent Candidates And Political Representation In India." *American Political Science Review* 112(3):678–697. Publisher: Cambridge University Press.

Kim, Jeong Hyun. 2019. "Direct Democracy And Women's Political Engagement." *American Journal of Political Science* 63(3):594–610. Publisher: Wiley Online Library.

Kocher, Matthew Adam, Thomas B. Pepinsky and Stathis N. Kalyvas. 2011. "Aerial Bombing Snd Counterinsurgency In The Vietnam War." *American Journal of Political Science* 55(2):201–218. Publisher: Wiley Online Library.

Kriner, Douglas L. and Eric Schickler. 2014. "Investigating The President: Committee Probes And Presidential Approval,1953Extendash2006." *The Journal of Politics* 76(2):521–534. Publisher: Cambridge University Press New York, USA.

Kuipers, Nicholas and Alexander Sahn. 2022. "The Representational Consequences of Municipal Civil Service Reform." *American Political Science Review* pp. 1–17.

Laitin, David D. and Rajesh Ramachandran. 2016. "Language Policy And Human Development." *American Political Science Review* 110(3):457–480. Publisher: Cambridge University Press.

Lei, Zhenhuan and Junlong Aaron Zhou. 2022. "Private returns to public investment: Political career incentives and infrastructure investment in China." *The Journal of Politics* 84(1):455–469.

Lelkes, Yphtach, Gaurav Sood and Shanto Iyengar. 2017. "The Hostile Audience: The Effect Of Access To Broadband Internet On Partisan Affect." *American Journal of Political Science* 61(1):5–20. Publisher: Wiley Online Library.

Leonardi, Robert, Raffaella Y Nanetti and Robert D Putnam. 2001. *Making Democracy Work: Civic Traditions In Modern Italy*. Princeton university press Princeton, NJ.

Lerman, Amy E., Meredith L. Sadin and Samuel Trachtman. 2017. "Policy Uptake As Political Behavior: Evidence From The Affordable Care Act." *The American Political Science Review* 111(4):755. Publisher: Cambridge University Press.

López-Moctezuma, Gabriel, Leonard Wantchekon, Daniel Rubenson, Thomas Fujiwara and Cecilia Pe Lero. 2022. "Policy Deliberation and Voter Persuasion: Experimental Evidence from an Election in the Philippines." *American Journal of Political Science* 66(1):59–74.

Lorentzen, Peter, Pierre Landry and John Yasuda. 2014. "Undermining Authoritarian Innovation: The Power Of China's Industrial Giants." *The Journal of Politics* 76(1):182–194. Publisher: Cambridge University Press New York, USA.

McClendon, Gwyneth H. 2014. "Social Esteem And Participation In Contentious Politics: A Field Experiment At An LGBT Pride Rally." *American Journal of Political Science* 58(2):279–290. Publisher: Wiley Online Library.

Meredith, Marc. 2013. "Exploiting Friends-and-neighbors To Estimate Coattail Effects." *American Political Science Review* p. 742–765. Publisher: JSTOR.

Nellis, Gareth and Niloufer Siddiqui. 2018. "Secular Party Rule And Religious Violence In Pakistan." *The American Political Science Review* 112(1):49. Publisher: Cambridge University Press.

Nunn, Nathan. 2008. "The Long-term Effects Of Africa's Slave Trades." *The Quarterly Journal of Economics* 123(1):139–176.

Nunn, Nathan and Leonard Wantchekon. 2011. "The Slave Trade And The Origins Of Mistrust In Africa." *American Economic Review* 101(7):3221–52.

Olea, José Luis Montiel and Carolin Pflueger. 2013. "A Robust Test For Weak Instruments." *Journal of business & economic statistics: a publication of the American Statistical Association* 31(3):358–369.

Pianzola, Joëlle, Alexander H. Trechsel, Kristjan Vassil, Guido Schwerdt and R. Michael Alvarez. 2019. "The Impact Of Personalized Information On Vote Intention: Evidence From A Randomized Field Experiment." *The Journal of Politics* 81(3):833–847. Publisher: The University of Chicago Press Chicago, IL.

Ritter, Emily Hencken and Courtenay R. Conrad. 2016. "Preventing And Responding To Dissent: The Observational Challenges Of Explaining Strategic Repression." *American Political Science Review* 110(1):85–99. Publisher: Cambridge University Press.

Rueda, Miguel R. 2017. "Small Aggregates, Big Manipulation: Vote Buying Enforcement And Collective Monitoring." *American Journal of Political Science* 61(1):163–177. Publisher: Wiley Online Library.

Schleiter, Petra and Margit Tavits. 2016. "The Electoral Benefits Of Opportunistic Election Timing." *The Journal of Politics* 78(3):836–850. Publisher: University of Chicago Press Chicago, IL.

Schubiger, Livia Isabella. 2021. "State violence and wartime civilian agency: Evidence from Peru." *The Journal of Politics* 83(4):1383–1398.

Sexton, Renard, Rachel L. Wellhausen and Michael G. Findley. 2019. "How Government Reactions To Violence Worsen Social Welfare: Evidence From Peru." *American Journal of Political Science* 63(2):353–367. Publisher: Wiley Online Library.

Spenkuch, Jörg L. and Philipp Tillmann. 2018. "Elite Influence? Religion And The Electoral Success Of The Nazis." *American Journal of Political Science* 62(1):19–36. Publisher: Wiley Online Library.

Stewart, Megan A. and Yu-Ming Liou. 2017. "Do Good Borders Make Good Rebels? Territorial Control And Civilian Casualties." *The Journal of Politics* 79(1):284–301. Publisher: University of Chicago Press Chicago, IL.

Stokes, Leah C. 2016. "Electoral Backlash Against Climate Policy: A Natural Experiment On Retrospective Voting And Local Resistance To Public Policy." *American Journal of Political Science* 60(4):958–974. Publisher: Wiley Online Library.

Tajima, Yuhki. 2013. "The Institutional Basis Of Intercommunal Order: Evidence From Indonesia's Democratic Transition." *American Journal of Political Science* 57(1):104–119. Publisher: Wiley Online Library.

Trounstine, Jessica. 2016. "Segregation And Inequality In Public Goods." *American Journal of Political Science* 60(3):709–725. Publisher: Wiley Online Library.

Urpelainen, Johannes and Alice Tianbo Zhang. 2022. "Electoral Backlash or Positive Reinforcement? Wind Power and Congressional Elections in the United States." *The Journal of Politics* 84(3):1306–1321.

Van Kippersluis, Hans and Cornelius A Rietveld. 2018. "Beyond Plausibly Exogenous." *The econometrics journal* 21(3):316–331.
URL: *https://academic.oup.com/ectj/article/21/3/316/5145983*

Vernby, Kare. 2013. "Inclusion And Public Policy: Evidence From Sweden's Introduction Of Noncitizen Suffrage." *American Journal of Political Science* 57(1):15–29.

Webster, Steven W, Elizabeth C Connors and Betsy Sinclair. 2022. "The social consequences of political anger." *The Journal of Politics* 84(3):1292–1305.

West, Emily A. 2017. "Descriptive Representation And Political Efficacy: Evidence From Obama And Clinton." *The Journal of Politics* 79(1):351–355. Publisher: University of Chicago Press Chicago, IL.

Wood, Abby K and Christian R Grose. 2022. "Campaign finance transparency affects legislators election outcomes and behavior." *American Journal of Political Science* 66(2):516–534.

Zhu, Boliang. 2017. "MNCs, Rents, And Corruption: Evidence From China." *American Journal of Political Science* 61(1):84–99.

Ziaja, Sebastian. 2020. "More Donors, More Democracy." *The Journal of Politics* 82(2):433–447. Publisher: The University of Chicago Press Chicago, IL.

Supplemental Materials
Appendix B

How Much Should We Trust Instrumental Variable Estimates in Political Science? Practical Advice based on 67 Replicated Studies

30 March 2023

Contents

Readme

- est_ols stores treatment effect estimates from the naive OLS estimation. 'Analytic' corresponds to analytic asymptotic standard errors (SEs) and confidence intervals (CIs). 'Boot.c' and 'Boot.t' represent inferential methods based on bootstrapped coefficients and bootstrapped t-statistics, respectively.
- est_2sls stores treatment effect estimates from the 2SLS estimation.
- AR stores results from the Anderson-Rubin test. The confidence region (CR) is produced by the inversion method. 'AR.bounded = TRUE' means that the CR is bounded and not empty.
- F.stat stores F statistics based on classic SEs (F.standard), H.W. robust SEs (F.robust), cluster-robust SEs (F.cluster), bootstrapped or cluster-bootstrapped SEs (F.bootstrap) and the effective F (F.effective). In the one-treatment-one-instrument case, F.effective is the same as F.robust (if there is no clustering structure) or F.cluster (if there is one).
- rho stores the partial correlation coefficient between the treatment and the predicted treatment from the first stage regression.
- tf.cF stores the results from the tF-cF procedure. Specifically, cF corresponds to the adjusted critical value based on the first stage (effective) F statistic for the subsequent t-test.
- est_rf stores the results from the reduced form regression. The control variables are partialled out.
- est_fs stores the results from the first stage regression. The control variables are partialled out.
- p_iv stores the number of instruments. N and N_cl stores the the number of observations and the number of clusters (if there is a clustering structure), respectively. df stores the degree of freedom from the 2SLS regression.
- nvalues stores the numbers of unique values in the outcome, treatment, and instrument.

APSR

Baccini and Weymouth (2021)

Replication Summary	
Unit of analysis	county
Treatment	Manufacturing Layoffs
Instrument	Bartik instrument
Outcome	Change of Democratic Vote Share
Model	Table2(3)

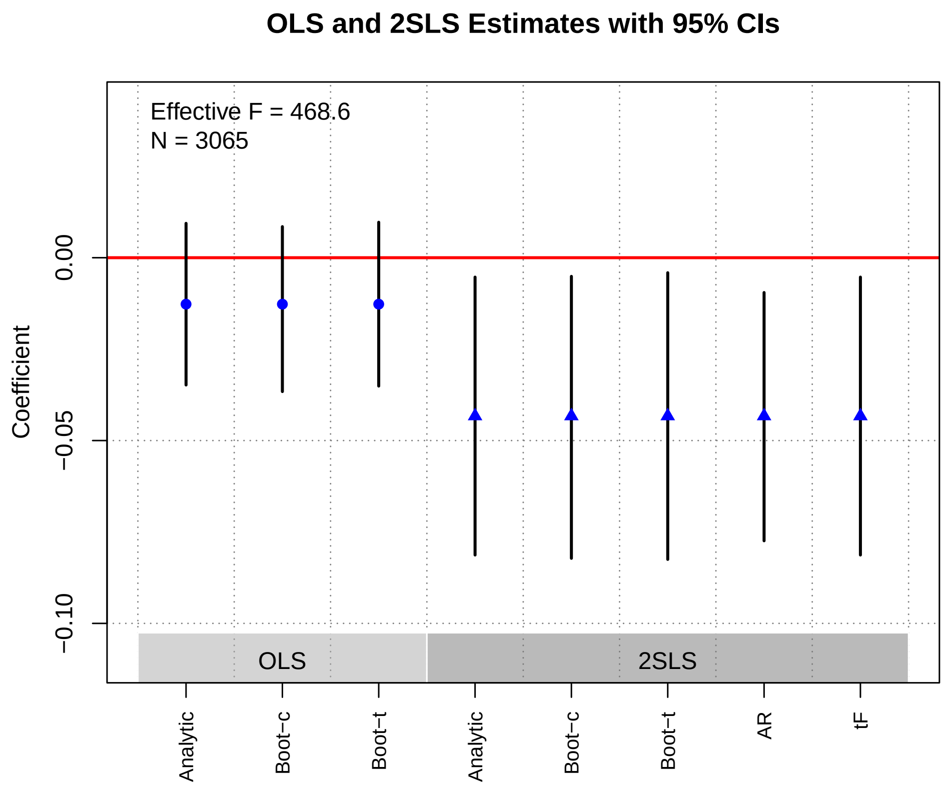
```
df <- readRDS("./data/apsr_baccini_etal_2021.rds")
D <-"msl_pc4y2"
Y <- "ddem_votes_pct1"
Z <-  "bartik_leo5"
controls <- c("LAU_unemp_rate_4y", "pers_m_total_share_4y", "pers_coll_share_4y",
              "white_counties_4y", "msl_service_pc4y")
cl <- NULL
FE <- "id_state"
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 30.221 sec.
## AR Test Inversion...

## $est_ols
##             Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -0.0127 0.0113 -1.1240 -0.0348   0.0094   0.261
## Boot.c   -0.0127 0.0115 -1.1059 -0.0366   0.0085   0.248
## Boot.t   -0.0127 0.0113 -1.1240 -0.0351   0.0097   0.278
##
## $est_2sls
##             Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -0.0433 0.0194 -2.2308 -0.0813  -0.0053  0.0257
## Boot.c   -0.0433 0.0198 -2.1822 -0.0822  -0.0051  0.0240
## Boot.t   -0.0433 0.0194 -2.2308 -0.0825  -0.0041  0.0290
##
## $AR
## $AR$Fstat
##         F       df1       df2         p
##    6.1879    1.0000 3063.0000    0.0129
##
```

```
## $AR$ci.print
## [1] "[-0.0774, -0.0095]"
##
## $AR$ci
## [1] -0.07740808 -0.00952044
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##   1537.5647    468.6180          NA    480.7918    468.6180
##
## $rho
## [1] 0.5815
##
## $tF
##        F       cF     Coef       SE        t   CI2.5%  CI97.5%  p-value
## 468.6180   1.9600  -0.0433   0.0194  -2.2308  -0.0813  -0.0053   0.0257
##
## $est_rf
##                Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## bartik_leo5 -4.5381 2.0355  0.0258 2.0617  -8.4894   -0.5509     0.024
##
## $est_fs
##                 Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## bartik_leo5 104.8786 4.8448       0 4.7831   95.739   114.338         0
##
## $p_iv
## [1] 1
##
## $N
## [1] 3065
##
## $N_cl
## NULL
##
## $df
## [1] 3010
##
## $nvalues
##      ddem_votes_pct1 msl_pc4y2 bartik_leo5
## [1,]            3062      2913        2771
```

```
plot_coef(g)
```



Blattman et al. (2014)

Replication Summary	
Unit of analysis	resident
Treatment	mass education campaign for dispute resolution
Instrument	assignment to treatment blocks
Outcome	serious land dispute
Model	Table9(8)

```
df <- readRDS("./data/apsr_Blattman_etal_2014.rds")
df$district <- 0
for (i in 1:15) {df$district[which(df[,paste0("district",i)]==1)] <- i}
D <-"months_treated"
Y <- "fightweap_dummy"
Z <- c("block1", "block2", "block3")
controls <- c("ageover60", "age40_60", "age20_40",
 "yrs_edu", "female", "stranger", "christian",
 "minority", "cashearn_imputedhst", "noland",
 "land_sizehst", "farm_sizehst", "lndtake_dum",
 "housetake_dum", "vsmall", "small",
 "small2", "small3", "quartdummy", "cedulevel_bc",
```

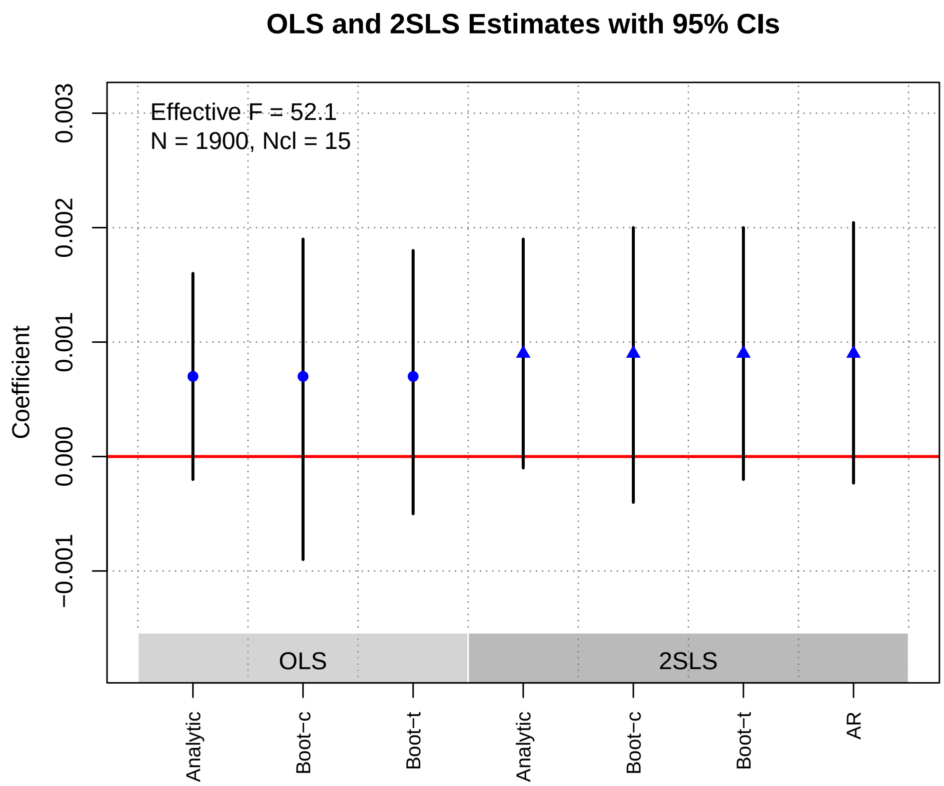
```r
 "ctownhh_log_el", "cwealthindex_bc", "cviol_experienced_bc",
 "clndtake_bc", "cviol_scale_bc", "clandconf_scale_bc",
 "cwitchcraft_scale_bc", "cpalaviol_imputed_bc",
 "cprog_ldr_beliefs_bc", "cattitudes_tribe_bc",
 "crelmarry_bc", "trainee")
cl <- "district"
FE <- "district"
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 32.677 sec.
## AR Test Inversion...

## $est_ols
##            Coef    SE      t CI 2.5% CI 97.5% p.value
## Analytic 7e-04 5e-04 1.4691  -2e-04   0.0016  0.1418
## Boot.c   7e-04 7e-04 0.9908  -9e-04   0.0019  0.3940
## Boot.t   7e-04 5e-04 1.4691  -5e-04   0.0018  0.2600
##
## $est_2sls
##            Coef    SE      t CI 2.5% CI 97.5% p.value
## Analytic 9e-04 5e-04 1.7837  -1e-04   0.0019  0.0745
## Boot.c   9e-04 6e-04 1.4913  -4e-04   0.0020  0.2260
## Boot.t   9e-04 5e-04 1.7837  -2e-04   0.0020  0.1020
##
## $AR
## $AR$Fstat
##         F       df1       df2         p
##    1.9496    3.0000 1896.0000    0.1196
##
## $AR$ci.print
## [1] "[-0.0002, 0.0020]"
##
## $AR$ci
## [1] -0.0002318028  0.0020441531
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##   F.standard    F.robust   F.cluster F.bootstrap F.effective
##    2756.3845   2472.2847    234.3492     98.6651     52.1000
##
```

```
## $rho
## [1] 0.9039
##
## $est_rf
##          Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## block1 0.0263 0.0123  0.0317 0.0134  -0.0045    0.0483     0.094
## block2 0.0027 0.0080  0.7312 0.0131  -0.0231    0.0283     0.874
## block3 0.0085 0.0086  0.3241 0.0108  -0.0163    0.0291     0.310
##
## $est_fs
##           Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## block1 20.0361 0.2680       0 1.2308  17.5315   22.5182     0.000
## block2 12.9786 0.2563       0 2.1291   9.0517   16.8681     0.000
## block3  6.7831 0.2513       0 1.8530   2.7445   10.1672     0.008
##
## $p_iv
## [1] 3
##
## $N
## [1] 1900
##
## $N_cl
## [1] 15
##
## $df
## [1] 14
##
## $nvalues
##      fightweap_dummy months_treated block1 block2 block3
## [1,]               2             34      2      2      2
```

```
plot_coef(g)
```



Colantone and Stanig (2018)

Replication Summary	
Unit of analysis	region
Treatment	regional-level import shock from China
Instrument	imports from China to the United States * local industrial structure
Outcome	leave share
Model	Table1(6)

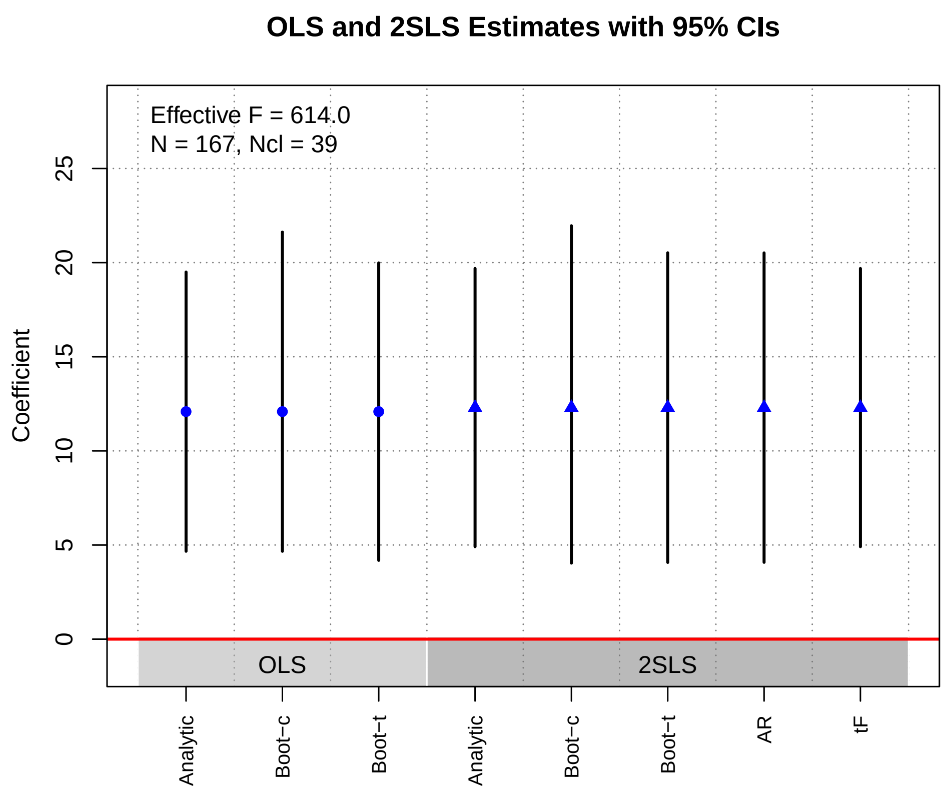
```
df<-readRDS("./data/apsr_Colantone_etal_2018.rds")
D <-'import_shock'
Y <- "leave_share"
Z <- "instrument_for_shock"
controls <- c("immigrant_share", "immigrant_arrivals")
cl <- "fix"
FE <- "nuts1"
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 28.552 sec.
## AR Test Inversion...
```

```
## $est_ols
##             Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 12.0854 3.7846 3.1933  4.6675  19.5033  0.0014
## Boot.c   12.0854 4.3379 2.7860  4.6668  21.6249  0.0020
## Boot.t   12.0854 3.7846 3.1933  4.1843  19.9865  0.0060
##
## $est_2sls
##             Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 12.2993 3.7701 3.2623  4.9099  19.6888  0.0011
## Boot.c   12.2993 4.5272 2.7168  4.0383  21.9601  0.0080
## Boot.t   12.2993 3.7701 3.2623  4.0754  20.5233  0.0070
##
## $AR
## $AR$Fstat
##        F      df1      df2        p
##   8.6843   1.0000 165.0000   0.0037
##
## $AR$ci.print
## [1] "[4.0805, 20.5182]"
##
## $AR$ci
## [1]  4.080471 20.518183
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##   2158.0662    792.4682    613.9804    608.4938    613.9804
##
## $rho
## [1] 0.9663
##
## $tF
##        F       cF     Coef       SE        t   CI2.5%  CI97.5%  p-value
## 613.9804   1.9600  12.2993   3.7701   3.2623   4.9099  19.6888   0.0011
##
## $est_rf
##                        Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## instrument_for_shock 1.5671 0.4798  0.0011 0.5829   0.5008    2.8488     0.008
##
## $est_fs
##                        Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## instrument_for_shock 0.1274 0.0045       0 0.0052   0.1186    0.1386         0
##
## $p_iv
```

```
## [1] 1
##
## $N
## [1] 167
##
## $N_cl
## [1] 39
##
## $df
## [1] 153
##
## $nvalues
##      leave_share import_shock instrument_for_shock
## [1,]         167          148                  148
```

```
plot_coef(g)
```



Croke et al. (2016)

Replication Summary	
Unit of analysis	individual
Treatment	education attainment
Instrument	access to the secondary education
Outcome	political participation

Replication Summary	
Model	Table2(b1)

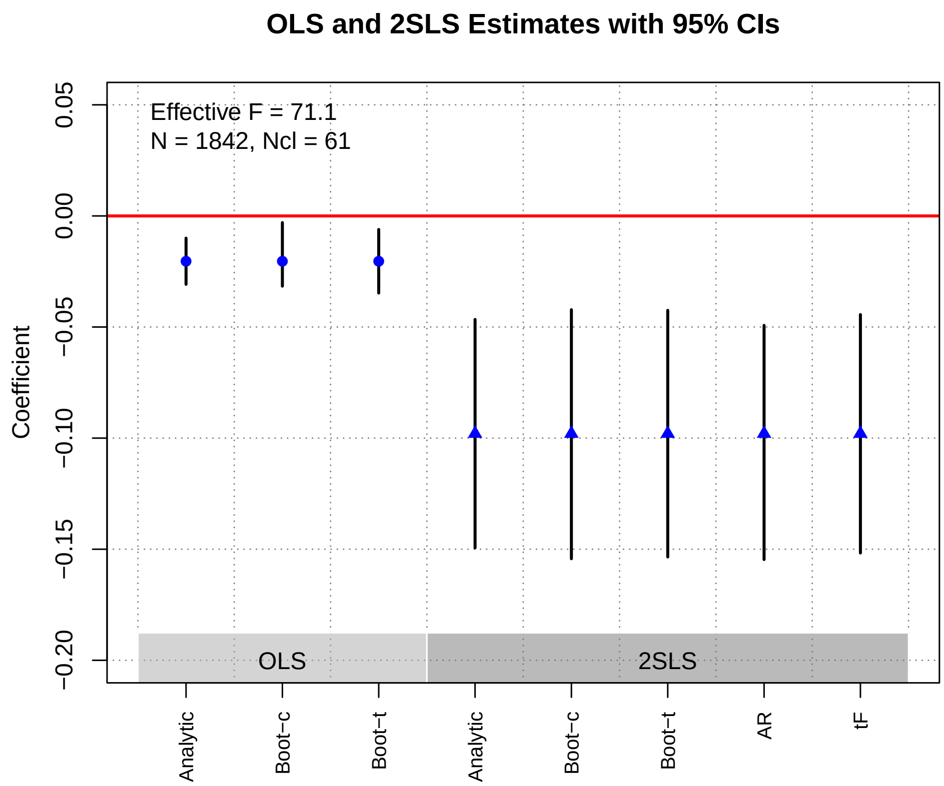
```
df <-readRDS("./data/apsr_Croke_etal_2016.rds")
D <- "edu"
Y <- "part_scale"
Z <- "treatment"
controls <-NULL
cl<- "district"
FE<- "year_survey"
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 29.095 sec.
## AR Test Inversion...

## $est_ols
##             Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -0.0204 0.0053 -3.8465 -0.0308  -0.0100  0.0001
## Boot.c   -0.0204 0.0077 -2.6429 -0.0316  -0.0030  0.0180
## Boot.t   -0.0204 0.0053 -3.8465 -0.0347  -0.0061  0.0060
##
## $est_2sls
##            Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -0.098 0.0262 -3.7385 -0.1494  -0.0466   2e-04
## Boot.c   -0.098 0.0279 -3.5176 -0.1543  -0.0422   0e+00
## Boot.t   -0.098 0.0262 -3.7385 -0.1535  -0.0425   0e+00
##
## $AR
## $AR$Fstat
##         F       df1       df2         p
##   15.6784    1.0000 1840.0000    0.0001
##
## $AR$ci.print
## [1] "[-0.1547, -0.0493]"
##
## $AR$ci
## [1] -0.15465266 -0.04925287
##
## $AR$bounded
## [1] TRUE
##
##
```

```
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##     79.7552     78.2588     71.1356     71.6651     71.1356
##
## $rho
## [1] 0.2041
##
## $tF
##       F      cF    Coef      SE       t  CI2.5% CI97.5% p-value
## 71.1356  2.0466 -0.0980  0.0262 -3.7385 -0.1517 -0.0444  0.0003
##
## $est_rf
##              Coef     SE p.value  SE.b CI.b2.5% CI.b97.5% p.value.b
## treatment -0.0657 0.0164   1e-04 0.017  -0.0971   -0.0297         0
##
## $est_fs
##             Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## treatment 0.6708 0.0758       0 0.0792   0.5265    0.8435         0
##
## $p_iv
## [1] 1
##
## $N
## [1] 1842
##
## $N_cl
## [1] 61
##
## $df
## [1] 1835
##
## $nvalues
##      part_scale edu treatment
## [1,]          7   7         5
```

```
plot_coef(g)
```



Dower et al. (2018) (a)

Replication Summary	
Unit of analysis	district*year
Treatment	frequency of unrest
Instrument	religious polarization
Outcome	peasant representation
Model	Table3(1)

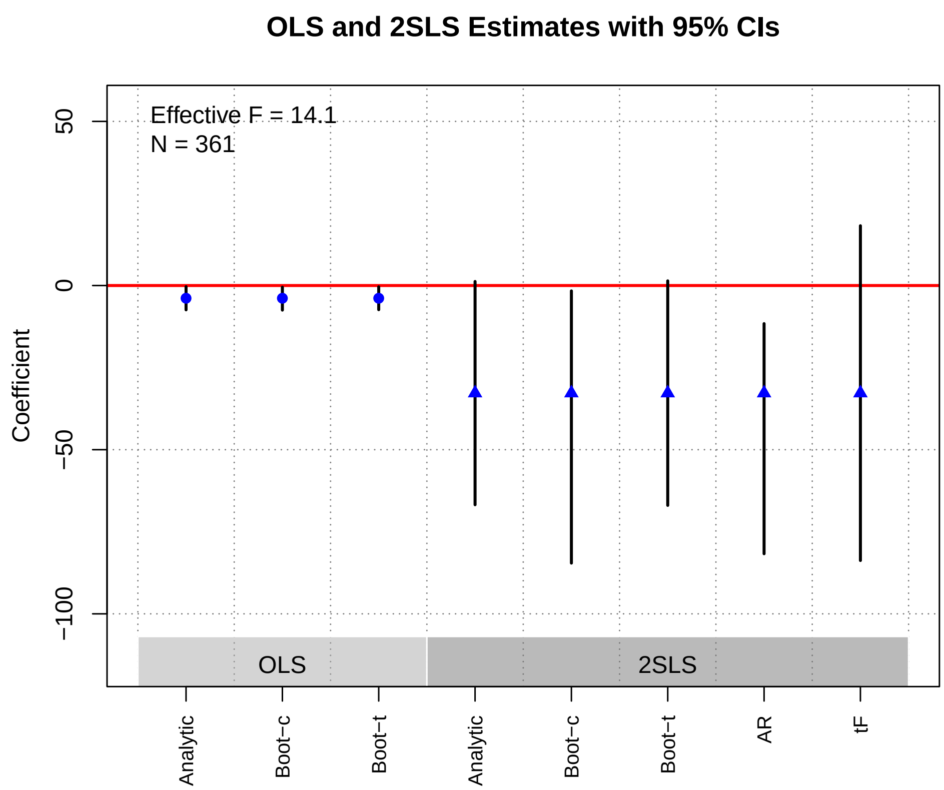
```
df <- readRDS("./data/apsr_Dower_etal_2018.rds")
D <-"afreq"
Y <-"peasantrepresentation_1864"
Z <-"religpolarf4_1870"
controls <- c("distance_moscow", "goodsoil", "lnurban", "lnpopn", "province_capital")
cl <- NULL
FE <- NULL
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 12.602 sec.
## AR Test Inversion...
```

```
## $est_ols
##             Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -3.8696 1.8013 -2.1483 -7.4001  -0.3391  0.0317
## Boot.c   -3.8696 1.7718 -2.1840 -7.4807  -0.5048  0.0220
## Boot.t   -3.8696 1.8013 -2.1483 -7.3572  -0.3820  0.0290
##
## $est_2sls
##              Coef      SE       t  CI 2.5% CI 97.5% p.value
## Analytic -32.7701 17.3518 -1.8886 -66.7796   1.2393  0.0589
## Boot.c   -32.7701 25.9097 -1.2648 -84.5533  -1.6303  0.0360
## Boot.t   -32.7701 17.3518 -1.8886 -66.9699   1.4297  0.0630
##
## $AR
## $AR$Fstat
##        F      df1      df2        p
##   9.4039   1.0000 359.0000   0.0023
##
## $AR$ci.print
## [1] "[-81.7021, -11.6010]"
##
## $AR$ci
## [1] -81.70209 -11.60096
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##     12.0237     14.0828          NA     13.7462     14.0828
##
## $rho
## [1] 0.1812
##
## $tF
##        F       cF     Coef       SE        t   CI2.5%  CI97.5%  p-value
##  14.0828   2.9384 -32.7701  17.3518  -1.8886 -83.7561  18.2158   0.2078
##
## $est_rf
##                     Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## religpolarf4_1870 -3.9279 1.8715  0.0358 1.8833  -7.4712   -0.1764     0.036
##
## $est_fs
##                    Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## religpolarf4_1870 0.1199 0.0319   2e-04 0.0323   0.0609    0.1823         0
##
## $p_iv
```

```
## [1] 1
##
## $N
## [1] 361
##
## $N_cl
## NULL
##
## $df
## [1] 354
##
## $nvalues
##      peasantrepresentation_1864 afreq religpolarf4_1870
## [1,]                        128    12               361
```

```
plot_coef(g)
```



Dower et al. (2018) (b)

Replication Summary	
Unit of analysis	district*year
Treatment	frequency of unrest
Instrument	religious polarization
Outcome	peasant representation

Replication Summary	
Model	Table1(2)

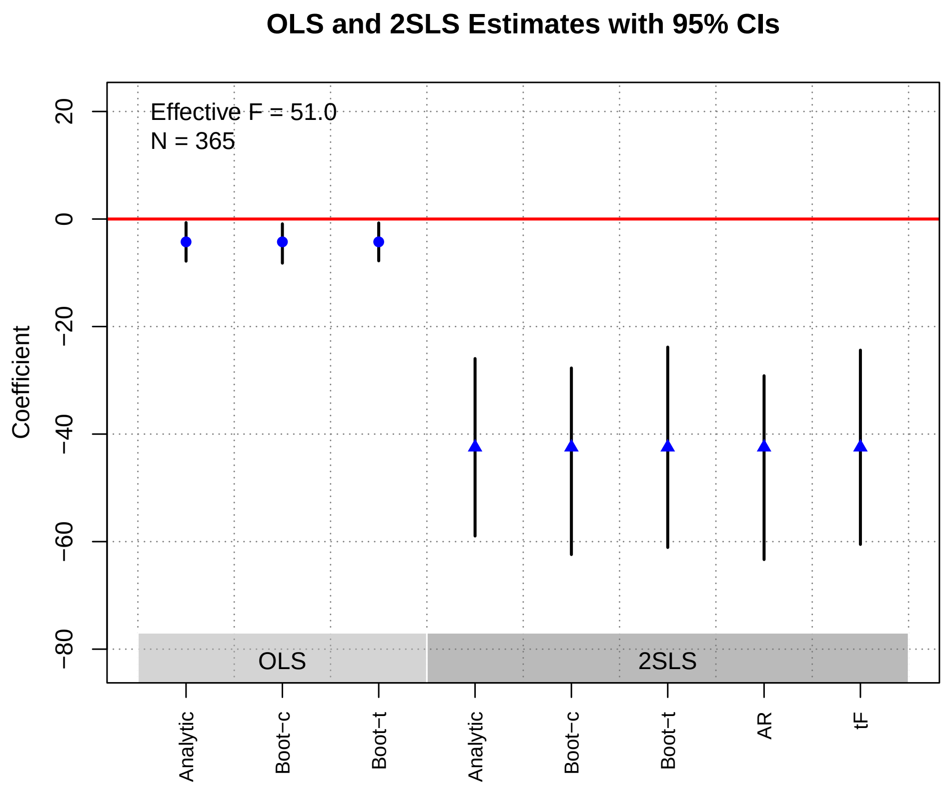
```r
df <- readRDS("./data/apsr_Dower_etal_2018.rds")
D <-"afreq"
Y <-"peasantrepresentation_1864"
Z <-"serfperc1"
controls <- c("distance_moscow", "goodsoil", "lnurban", "lnpopn", "province_capital")
cl <- NULL
FE <- NULL
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 12.597 sec.
## AR Test Inversion...
```

```
## $est_ols
##             Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -4.2492 1.8297 -2.3224 -7.8353  -0.6631  0.0202
## Boot.c   -4.2492 1.8488 -2.2984 -8.2011  -0.8969  0.0120
## Boot.t   -4.2492 1.8297 -2.3224 -7.7727  -0.7258  0.0220
##
## $est_2sls
##              Coef     SE       t  CI 2.5% CI 97.5% p.value
## Analytic -42.4545 8.4195 -5.0424 -58.9567 -25.9522   0.000
## Boot.c   -42.4545 9.0521 -4.6900 -62.4018 -27.7095   0.000
## Boot.t   -42.4545 8.4195 -5.0424 -61.0869 -23.8220   0.001
##
## $AR
## $AR$Fstat
##        F      df1      df2        p
##  63.9521   1.0000 363.0000   0.0000
##
## $AR$ci.print
## [1] "[-63.3348, -29.1517]"
##
## $AR$ci
## [1] -63.33480 -29.15166
##
## $AR$bounded
## [1] TRUE
##
##
```

```
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##     47.6256     51.0176          NA     51.0622     51.0176
##
## $rho
## [1] 0.3427
##
## $tF
##        F       cF     Coef       SE        t   CI2.5%  CI97.5%  p-value
##  51.0176   2.1457 -42.4545   8.4195  -5.0424 -60.5204 -24.3885   0.0000
##
## $est_rf
##               Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## serfperc1 -11.7823 1.6414       0 1.6722 -15.0422    -8.572         0
##
## $est_fs
##             Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## serfperc1 0.2775 0.0389       0 0.0388   0.2025    0.3561         0
##
## $p_iv
## [1] 1
##
## $N
## [1] 365
##
## $N_cl
## NULL
##
## $df
## [1] 358
##
## $nvalues
##      peasantrepresentation_1864 afreq serfperc1
## [1,]                        128    12       361
```

```
plot_coef(g)
```



Gerber et al. (2010)

Replication Summary	
Unit of analysis	individual
Treatment	aligning party identification with latent partisanship
Instrument	being sent mail
Outcome	voting and party alignment scale
Model	Table4(1)

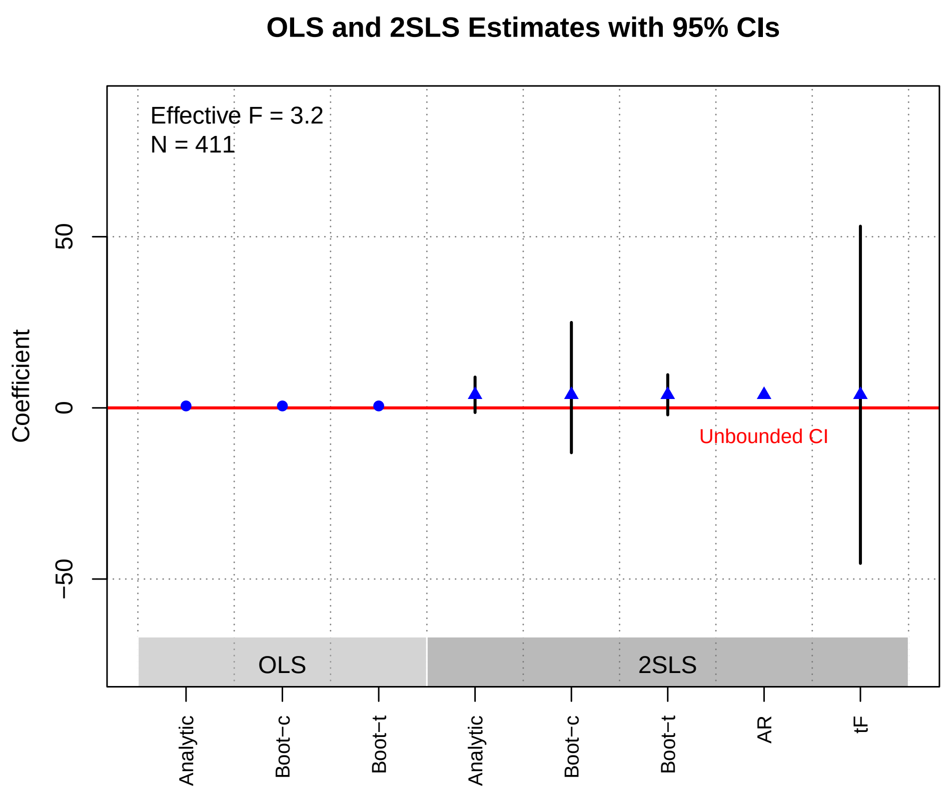
```
df <- readRDS("./data/apsr_Gerber_etal_2010.rds")
D <-"pt_id_with_lean"
Y <- "pt_voteevalalignindex"
Z <- "treat"
controls <- c("pre_lean_dem", "age", "age2" ,"regyear" ,
              "regyearmissing", "twonames", "combined_female",
              "voted2006", "voted2004", "voted2002", "voted2000",
              "voted1998", "voted1996", "interest", "pre_aligned_vh",
              "pre_direct_unemp", "pre_direct_econ","pre_direct_bushap",
              "pre_direct_congapp")
cl <- NULL
FE <- NULL
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 12.970 sec.
## AR Test Inversion...

## $est_ols
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.5658 0.1709 3.3105  0.2308   0.9008   9e-04
## Boot.c   0.5658 0.1761 3.2123  0.2250   0.9089   2e-03
## Boot.t   0.5658 0.1709 3.3105  0.2198   0.9117   3e-03
##
## $est_2sls
##            Coef      SE      t  CI 2.5% CI 97.5% p.value
## Analytic 3.8231  2.6392 1.4486  -1.3497   8.9960  0.1475
## Boot.c   3.8231 14.1908 0.2694 -13.0997  24.9545  0.1320
## Boot.t   3.8231  2.6392 1.4486  -2.0424   9.6887  0.1400
##
## $AR
## $AR$Fstat
##        F      df1      df2        p
##   3.9122   1.0000 409.0000   0.0486
##
## $AR$ci.print
## [1] "[0.0754, Inf)"
##
## $AR$ci
## [1] 0.07543774        Inf
##
## $AR$bounded
## [1] FALSE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##      2.9926      3.1563          NA      3.0871      3.1563
##
## $rho
## [1] 0.0873
##
## $tF
##        F       cF     Coef       SE        t   CI2.5%  CI97.5%  p-value
##   3.1563  18.6600   3.8231   2.6392   1.4486 -45.4249  53.0712   0.8791
##
## $est_rf
##         Coef     SE p.value SE.b CI.b2.5% CI.b97.5% p.value.b
## treat 0.2742 0.1429  0.0551 0.14  -0.0102     0.538      0.06
##
```

```
## $est_fs
##         Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## treat 0.0717 0.0404  0.0756 0.0408  -0.0065    0.1558     0.074
##
## $p_iv
## [1] 1
##
## $N
## [1] 411
##
## $N_cl
## NULL
##
## $df
## [1] 390
##
## $nvalues
##      pt_voteevalalignindex pt_id_with_lean treat
## [1,]                    10               2     2
```

```
plot_coef(g)
```



Hager et al. (2019)

Replication Summary	
Unit of analysis	individual
Treatment	ethnic riots (destruction)
Instrument	distance to the nearest location where armored military vehicles were stolen
Outcome	prosocial behavior
Model	Figure6

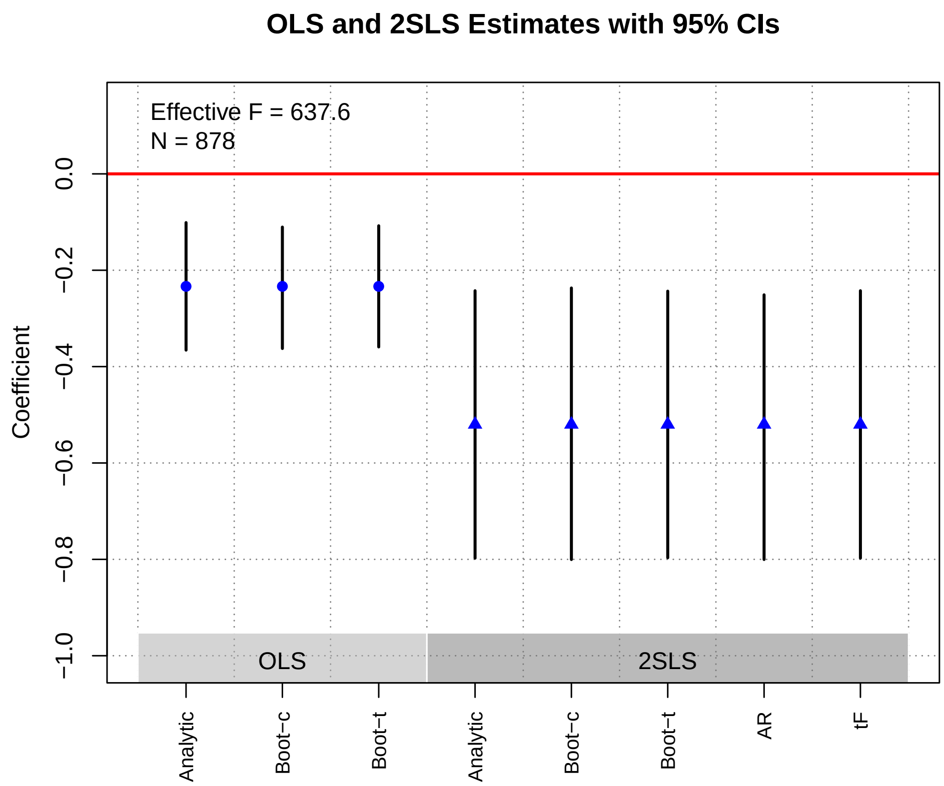
```
df <- readRDS("./data/apsr_Hager_etal_2019.rds")
D <-"affected"
Y <- "pd_in_scale"
Z <-  "apc_min_distance"
controls <- NULL
cl <- NULL
FE <- NULL
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 12.503 sec.
## AR Test Inversion...

## $est_ols
##              Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -0.2335 0.0675 -3.4582 -0.3658  -0.1011   5e-04
## Boot.c   -0.2335 0.0664 -3.5158 -0.3624  -0.1104   0e+00
## Boot.t   -0.2335 0.0675 -3.4582 -0.3592  -0.1077   0e+00
##
## $est_2sls
##            Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -0.52 0.1416 -3.6733 -0.7975  -0.2425   2e-04
## Boot.c   -0.52 0.1435 -3.6244 -0.8004  -0.2368   0e+00
## Boot.t   -0.52 0.1416 -3.6733 -0.7967  -0.2434   0e+00
##
## $AR
## $AR$Fstat
##        F      df1      df2        p
##  14.2026   1.0000 876.0000   0.0002
##
## $AR$ci.print
## [1] "[-0.8003, -0.2510]"
##
## $AR$ci
## [1] -0.8003132 -0.2510336
##
```

```
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##    271.8565    637.5699          NA    578.5529    637.5699
##
## $rho
## [1] 0.4867
##
## $tF
##        F       cF     Coef       SE        t   CI2.5%  CI97.5%  p-value
## 637.5699   1.9600  -0.5200   0.1416  -3.6733  -0.7975  -0.2425   0.0002
##
## $est_rf
##                   Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## apc_min_distance 0.1011 0.0272   2e-04 0.0276   0.0466    0.1529         0
##
## $est_fs
##                    Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## apc_min_distance -0.1943 0.0077       0 0.0081   -0.209   -0.1779         0
##
## $p_iv
## [1] 1
##
## $N
## [1] 878
##
## $N_cl
## NULL
##
## $df
## [1] 876
##
## $nvalues
##      pd_in_scale affected apc_min_distance
## [1,]           2        2              193
```

```
plot_coef(g)
```



Hager and Krakowski (2022)

Replication Summary	
Unit of analysis	individual
Treatment	number of secret police officers
Instrument	number of corrupted Catholic priests
Outcome	resistance
Model	Table3(2)

```
df <- readRDS("./data/apsr_Hager_Krakowski_2022.rds")

D <-"commanders"
Y <- "y"
Z <-  "priests_continuous"
controls <- NULL
cl <- NULL
FE <- NULL
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

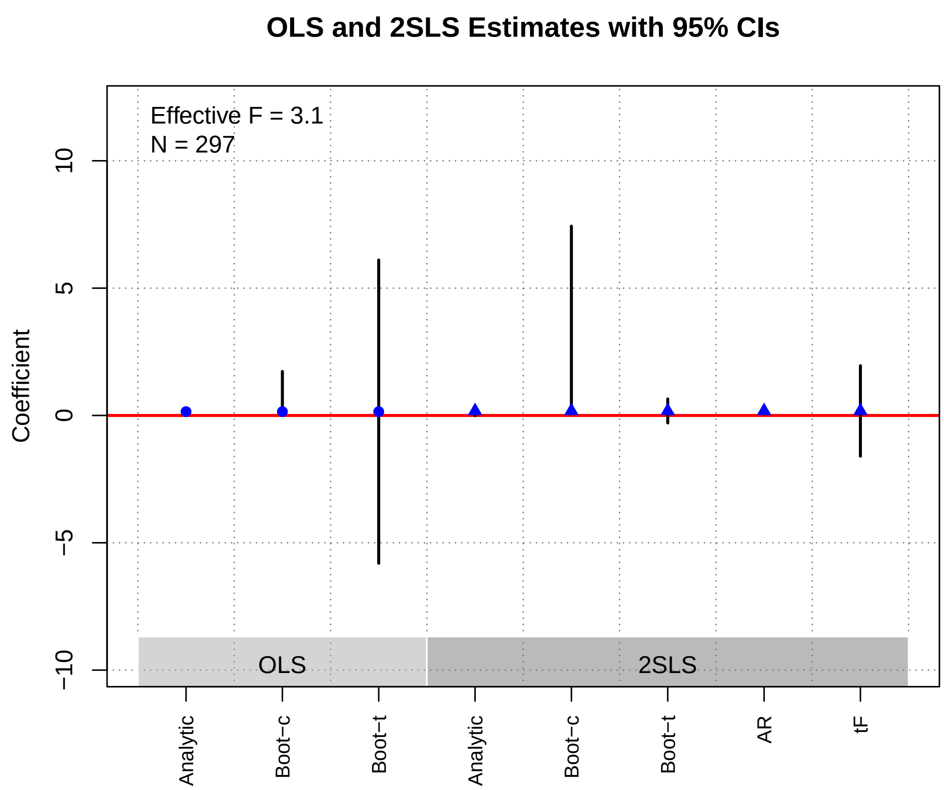
```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 12.587 sec.
```

```
## AR Test Inversion...

## $est_ols
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.1494 0.0751 1.9891  0.0022   0.2965  0.0467
## Boot.c   0.1494 0.3774 0.3957  0.0593   1.7277  0.0000
## Boot.t   0.1494 0.0751 1.9891 -5.8062   6.1049  0.5020
##
## $est_2sls
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.1765 0.0952 1.8537 -0.0101   0.3632  0.0638
## Boot.c   0.1765 2.5430 0.0694  0.0818   7.4345  0.0000
## Boot.t   0.1765 0.0952 1.8537 -0.2949   0.6479  0.3490
##
## $AR
## $AR$Fstat
##        F      df1      df2        p
##  52.7662   1.0000 295.0000   0.0000
##
## $AR$ci.print
## [1] "[0.1384, 0.2184]"
##
## $AR$ci
## [1] 0.1384381 0.2184313
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##    109.0543      3.1403          NA      3.1970      3.1403
##
## $rho
## [1] 0.5195
##
## $tF
##       F      cF    Coef      SE       t  CI2.5% CI97.5% p-value
##  3.1403 18.6600  0.1765  0.0952  1.8537 -1.6005  1.9535  0.8456
##
## $est_rf
##                      Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## priests_continuous 0.4736 0.1603  0.0031 0.1749   0.1753    0.8827         0
##
## $est_fs
##                      Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## priests_continuous 2.6827 1.5139  0.0764 1.5004   0.0229    5.3551         0
```

```
##
## $p_iv
## [1] 1
##
## $N
## [1] 297
##
## $N_cl
## NULL
##
## $df
## [1] 295
##
## $nvalues
##        y commanders priests_continuous
## [1,] 14         12                  7
```

```
plot_coef(g)
```



Kapoor and Magesan (2018)

Replication Summary	
Unit of analysis	constituency*election
Treatment	number of independent candidates

Replication Summary	
Instrument	changes in entry costs
Outcome	voter turnout
Model	Table4(b5)

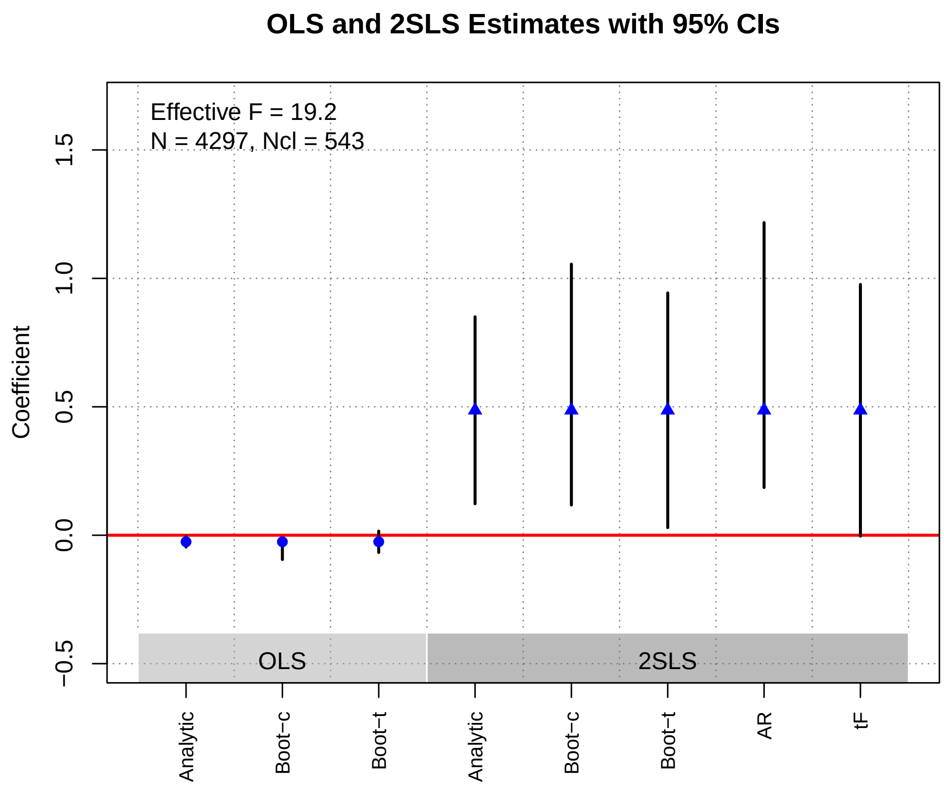
```
df<-readRDS("./data/apsr_Kapoor_etal_2018.rds")
D <-'CitCand'
Y <- "Turnout"
Z <- "UnScheduledDepChange"
controls <- c("CitCandBaseTrend", "CitCandBaseTrendSq", "CitCandBaseTrendCu",
              "CitCandBaseTrendQu",  "TurnoutBaseTrend", "TurnoutBaseTrendSq",
              "TurnoutBaseTrendCu", "TurnoutBaseTrendQu", "LnElectors",
              "LagWinDist", "LagWinDistSq", "LagWinDistCu",
              "LagWinDistQu", "LagTightElection")
cl<- "constituency"
FE <- c("year","constituency")
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 52.314 sec.
## AR Test Inversion...

## $est_ols
##             Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -0.0256 0.0105 -2.4375 -0.0462  -0.0050  0.0148
## Boot.c   -0.0256 0.0207 -1.2383 -0.0947  -0.0131  0.0000
## Boot.t   -0.0256 0.0105 -2.4375 -0.0672   0.0160  0.1510
##
## $est_2sls
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.4864 0.1856 2.6200  0.1225   0.8503  0.0088
## Boot.c   0.4864 0.2629 1.8502  0.1175   1.0555  0.0000
## Boot.t   0.4864 0.1856 2.6200  0.0296   0.9432  0.0370
##
## $AR
## $AR$Fstat
##         F       df1       df2         p
##   11.6079    1.0000 4295.0000    0.0007
##
## $AR$ci.print
## [1] "[0.1856, 1.2172]"
##
## $AR$ci
```

```
## [1] 0.1856426 1.2171638
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##     11.2301     23.7168     19.1635     18.9084     19.1635
##
## $rho
## [1] 0.0548
##
## $tF
##       F      cF    Coef      SE       t  CI2.5% CI97.5% p-value
## 19.1635  2.6390  0.4864  0.1856  2.6200 -0.0035  0.9763  0.0517
##
## $est_rf
##                        Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## UnScheduledDepChange -1.277 0.3929  0.0012 0.4461  -2.1331   -0.3284         0
##
## $est_fs
##                         Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## UnScheduledDepChange -2.6256 0.5391       0 0.6038   -3.914   -1.5902         0
##
## $p_iv
## [1] 1
##
## $N
## [1] 4297
##
## $N_cl
## [1] 543
##
## $df
## [1] 542
##
## $nvalues
##      Turnout CitCand UnScheduledDepChange
## [1,]    4293      68                    2
```

```
plot_coef(g)
```



Kuipers and Sahn (2022)

Replication Summary	
Unit of analysis	municipality* year
Treatment	civil service reform
Instrument	statewide assignment mandate
Outcome	descriptive representation on an unrestricted sample
Model	Table1(2)

```
df <- readRDS("./data/apsr_kuipers_2022.rds")
df<-df%>%filter(occ=='blue_collar' & name=='white_x_native_born')
D <-"treat_actual"
Y <- "govt"
Z <-  "treat_assign"
controls <-"pop"
cl <- NULL
FE <- c("YEAR","city")
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

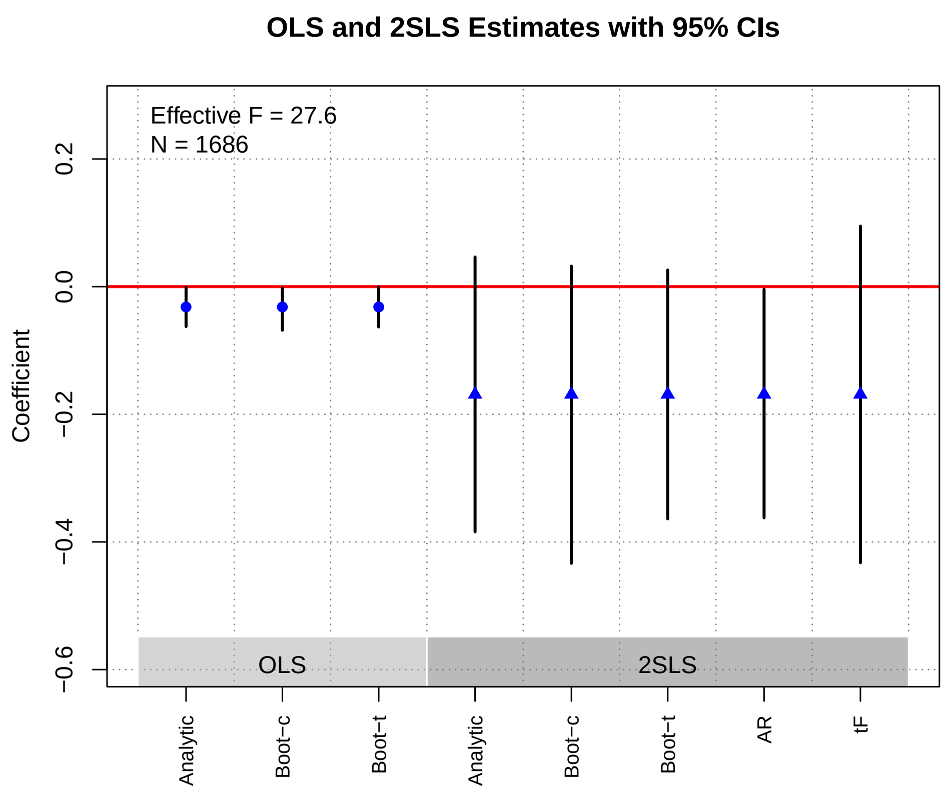
```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 30.603 sec.
```

```
## AR Test Inversion...

## $est_ols
##             Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -0.0319 0.0156 -2.0467 -0.0625  -0.0014  0.0407
## Boot.c   -0.0319 0.0168 -1.9037 -0.0684  -0.0025  0.0360
## Boot.t   -0.0319 0.0156 -2.0467 -0.0632  -0.0006  0.0480
##
## $est_2sls
##             Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -0.1689 0.1099 -1.5373 -0.3842   0.0464  0.1242
## Boot.c   -0.1689 0.1194 -1.4145 -0.4335   0.0321  0.1040
## Boot.t   -0.1689 0.1099 -1.5373 -0.3638   0.0260  0.0890
##
## $AR
## $AR$Fstat
##         F       df1       df2         p
##    4.0066    1.0000 1684.0000    0.0455
##
## $AR$ci.print
## [1] "[-0.3623, -0.0041]"
##
## $AR$ci
## [1] -0.362254764 -0.004101059
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##     32.4157     27.5670          NA     24.2805     27.5670
##
## $rho
## [1] 0.153
##
## $tF
##       F      cF    Coef      SE       t  CI2.5% CI97.5% p-value
## 27.5670  2.3999 -0.1689  0.1099 -1.5373 -0.4326  0.0948  0.2093
##
## $est_rf
##                 Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## treat_assign -0.0254 0.0162   0.116 0.0172  -0.0603    0.0054     0.104
##
## $est_fs
##                Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## treat_assign 0.1504 0.0286       0 0.0305   0.1004    0.2213         0
```

```
##
## $p_iv
## [1] 1
##
## $N
## [1] 1686
##
## $N_cl
## NULL
##
## $df
## [1] 1352
##
## $nvalues
##      govt treat_actual treat_assign
## [1,]  658            2            2
```

```
plot_coef(g)
```



Laitin and Ramachandran (2016)

Replication Summary	
Unit of analysis	country
Treatment	language choice

Replication Summary	
Instrument	geographic distance from the origins of writing
Outcome	human development index
Model	Table10(10)

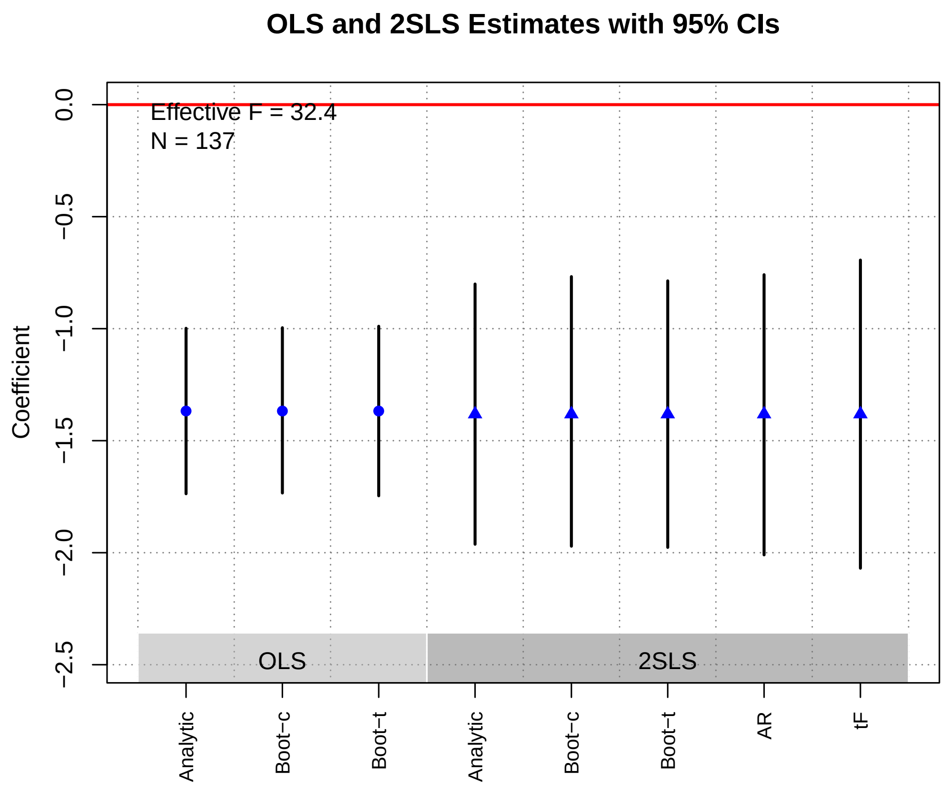
```
df <-readRDS("./data/apsr_Laitin_2016.rds")
D <-"avgdistance_delta50"
Y <- "zhdi_2010"
Z <- "DIST_BGNC"
controls <- c("cdf2003","ln_GDP_Indp", "edes1975",
              "America","xconst")
cl<- NULL
FE<- NULL
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 12.637 sec.
## AR Test Inversion...
```

```
## $est_ols
##             Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -1.3676 0.1884 -7.2594 -1.7369  -0.9984       0
## Boot.c   -1.3676 0.1861 -7.3491 -1.7335  -0.9958       0
## Boot.t   -1.3676 0.1884 -7.2594 -1.7460  -0.9893       0
##
## $est_2sls
##             Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -1.3815 0.2963 -4.6618 -1.9623  -0.8007       0
## Boot.c   -1.3815 0.3203 -4.3127 -1.9710  -0.7677       0
## Boot.t   -1.3815 0.2963 -4.6618 -1.9763  -0.7867       0
##
## $AR
## $AR$Fstat
##        F      df1      df2        p
##  15.0853   1.0000 135.0000   0.0002
##
## $AR$ci.print
## [1] "[-2.0097, -0.7592]"
##
## $AR$ci
## [1] -2.0097489 -0.7591826
##
## $AR$bounded
```

```
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##     55.1871     32.4040          NA     32.6764     32.4040
##
## $rho
## [1] 0.5459
##
## $tF
##       F      cF    Coef      SE       t  CI2.5% CI97.5% p-value
## 32.4040  2.3208 -1.3815  0.2963 -4.6618 -2.0692 -0.6938  0.0001
##
## $est_rf
##             Coef SE p.value SE.b CI.b2.5% CI.b97.5% p.value.b
## DIST_BGNC -1e-04  0   9e-04    0   -2e-04         0         0
##
## $est_fs
##            Coef SE p.value SE.b CI.b2.5% CI.b97.5% p.value.b
## DIST_BGNC 1e-04  0       0    0    1e-04     1e-04         0
##
## $p_iv
## [1] 1
##
## $N
## [1] 137
##
## $N_cl
## NULL
##
## $df
## [1] 130
##
## $nvalues
##      zhdi_2010 avgdistance_delta50 DIST_BGNC
## [1,]       121                  93       134
```

```
plot_coef(g)
```



Meredith (2013)

Replication Summary	
Unit of analysis	down-ballot race
Treatment	Democratic governor
Instrument	governor's home county
Outcome	down-ballot Democratic candidates' vote share
Model	Table3(5)

```
df <-readRDS("./data/apsr_Meredith_2013.rds")
Y <- "DemShareDB_res"
D<-"DemShareGOV_res"
Z <- "HomeGOV_res"
controls <- "HomeDB_res"
cl <- "fips"
FE<- NULL
weights<-NULL
(g <- ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

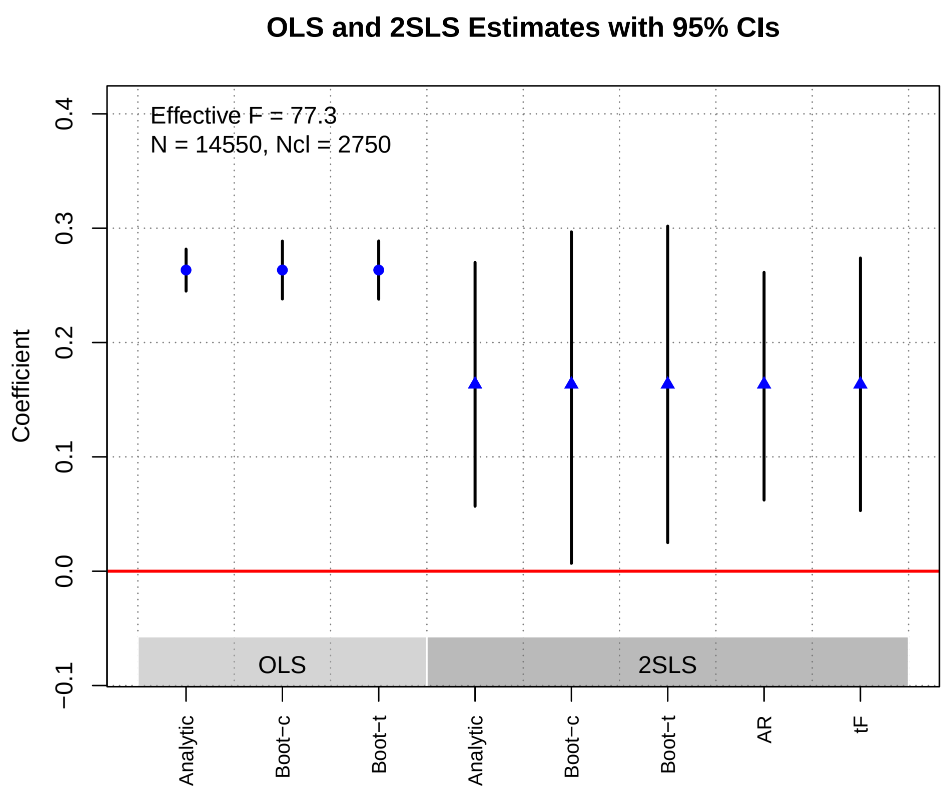
```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 17.872 sec.
## AR Test Inversion...
```

```
## Parallelising on 15 cores

## $est_ols
##            Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic 0.2634 0.0094 28.0999  0.2450   0.2817       0
## Boot.c   0.2634 0.0129 20.3403  0.2381   0.2887       0
## Boot.t   0.2634 0.0094 28.0999  0.2379   0.2888       0
##
## $est_2sls
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.1634 0.0544 3.0036  0.0568   0.2701  0.0027
## Boot.c   0.1634 0.0732 2.2315  0.0069   0.2968  0.0360
## Boot.t   0.1634 0.0544 3.0036  0.0250   0.3018  0.0170
##
## $AR
## $AR$Fstat
##          F        df1        df2          p
##     9.6035     1.0000 14548.0000     0.0019
##
## $AR$ci.print
## [1] "[0.0622, 0.2614]"
##
## $AR$ci
## [1] 0.0622274 0.2613756
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##    284.9652    141.9189     77.2953     75.2220     77.2953
##
## $rho
## [1] 0.1386
##
## $tF
##       F      cF    Coef      SE       t  CI2.5% CI97.5% p-value
## 77.2953  2.0300  0.1634  0.0544  3.0036  0.0530  0.2739  0.0037
##
## $est_rf
##               Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## HomeGOV_res 0.0062 0.0022  0.0052 0.0029    2e-04    0.0117     0.036
##
## $est_fs
##               Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## HomeGOV_res 0.0379 0.0032       0 0.0044   0.0295    0.0473         0
```

```
##
## $p_iv
## [1] 1
##
## $N
## [1] 14550
##
## $N_cl
## [1] 2750
##
## $df
## [1] 14547
##
## $nvalues
##      DemShareDB_res DemShareGOV_res HomeGOV_res
## [1,]          14550           14550        1466
```

```
plot_coef(g)
```



Nellis and Siddiqui (2018)

Replication Summary	
Unit of analysis	district*election

Replication Summary	
Treatment	the proportion of MNA seats in a district won by secularist candidates
Instrument	narrow victory by secular parties in a district
Outcome	religious violence
Model	Table2(1)

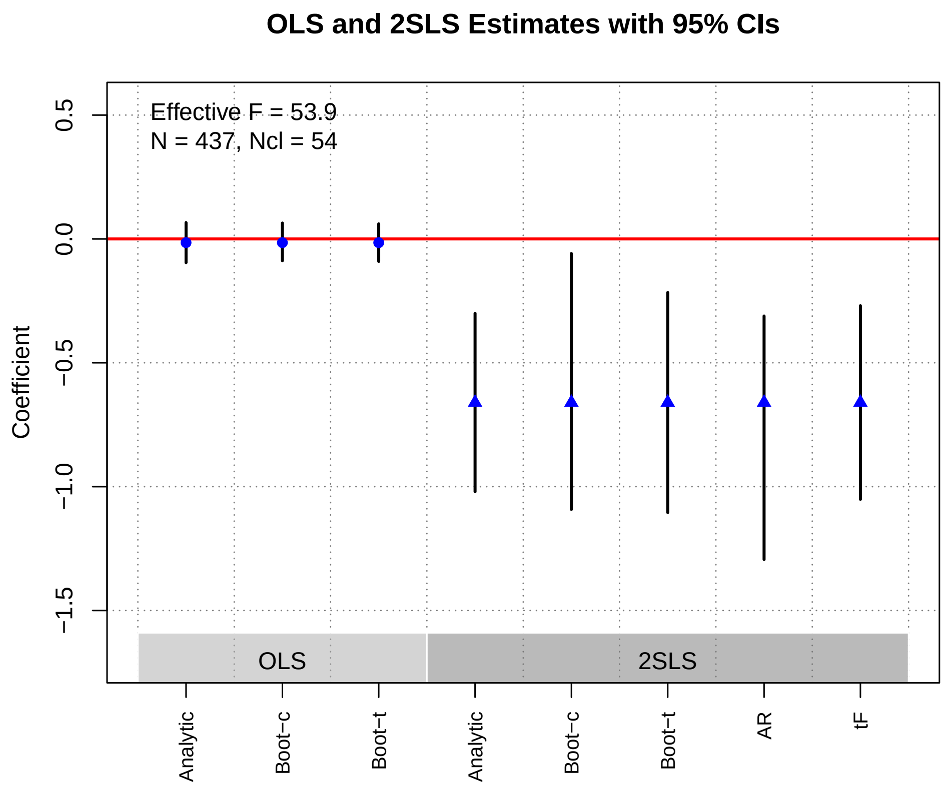
```r
df<-readRDS("./data/apsr_Nellis_etal_2018.rds")
D <-'secular_win'
Y <- "any_violence"
Z <- "secular_close_win"
controls <-"secular_close_race"
cl <- "cluster_var"
FE <- "pro"
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 31.082 sec.
## AR Test Inversion...

## $est_ols
##             Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -0.015 0.0413 -0.3620 -0.0959   0.0660  0.7174
## Boot.c   -0.015 0.0385 -0.3882 -0.0878   0.0643  0.6960
## Boot.t   -0.015 0.0413 -0.3620 -0.0908   0.0609  0.6870
##
## $est_2sls
##             Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -0.6603 0.1838 -3.5924 -1.0206  -0.3001  0.0003
## Boot.c   -0.6603 0.2558 -2.5811 -1.0918  -0.0590  0.0360
## Boot.t   -0.6603 0.1838 -3.5924 -1.1044  -0.2163  0.0140
##
## $AR
## $AR$Fstat
##       F      df1      df2        p
##  16.0695   1.0000 435.0000   0.0001
##
## $AR$ci.print
## [1] "[-1.2942, -0.3111]"
##
## $AR$ci
## [1] -1.2942296 -0.3110881
##
```

```
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##     22.0208     60.0400     53.9103     41.7554     53.9103
##
## $rho
## [1] 0.2207
##
## $tF
##       F      cF    Coef      SE       t  CI2.5% CI97.5% p-value
## 53.9103  2.1258 -0.6603  0.1838 -3.5924 -1.0511 -0.2696  0.0009
##
## $est_rf
##                     Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## secular_close_win -0.5965 0.1499   1e-04 0.2028  -0.8685    -0.058     0.036
##
## $est_fs
##                    Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## secular_close_win 0.9034 0.1166       0 0.1398    0.627    1.1879         0
##
## $p_iv
## [1] 1
##
## $N
## [1] 437
##
## $N_cl
## [1] 54
##
## $df
## [1] 430
##
## $nvalues
##      any_violence secular_win secular_close_win
## [1,]            2          26                17
```

```
plot_coef(g)
```



Ritter and Conrad (2016)

Replication Summary	
Unit of analysis	province in 54 African countries*day
Treatment	mobilized dissent
Instrument	rainfall
Outcome	repression
Model	Table1(3b)

```
df <- readRDS("./data/apsr_Ritter_etal_2016.rds")
D <- "dissentcount"
Y <- "represscount"
Z <- c("lograin", "rainannualpct")
controls <-"urban_mean"
cl<- NULL
FE<- NULL
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

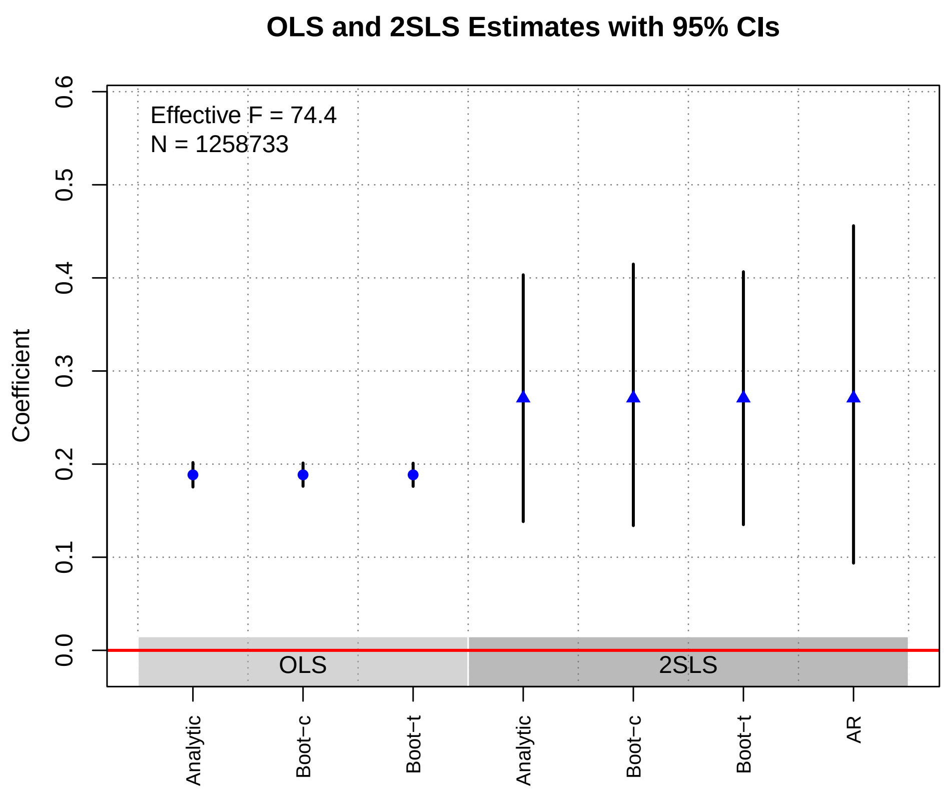
```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 5.297 sec.
## AR Test Inversion...
```

```
## Parallelising on 15 cores

## $est_ols
##            Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic 0.1885 0.0067 28.0525  0.1754   0.2017       0
## Boot.c   0.1885 0.0064 29.6088  0.1762   0.2012       0
## Boot.t   0.1885 0.0067 28.0525  0.1761   0.2010       0
##
## $est_2sls
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.2708 0.0676 4.0058  0.1383   0.4033   1e-04
## Boot.c   0.2708 0.0696 3.8902  0.1340   0.4148   0e+00
## Boot.t   0.2708 0.0676 4.0058  0.1350   0.4066   0e+00
##
## $AR
## $AR$Fstat
##           F         df1         df2           p
## 6.59790e+00 2.00000e+00 1.25873e+06 1.40000e-03
##
## $AR$ci.print
## [1] "[0.0937, 0.4560]"
##
## $AR$ci
## [1] 0.09367711 0.45601295
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##     58.3505     73.6819          NA     73.1541     74.3587
##
## $rho
## [1] 0.0096
##
## $est_rf
##                  Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## lograin        0.0001 0.0000  0.0000 0.0000    1e-04    0.0002     0.000
## rainannualpct -0.0092 0.0059  0.1199 0.0058   -2e-02    0.0028     0.122
##
## $est_fs
##                  Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## lograin        0.0005 0.0000   0e+00 0.0000   0.0004    0.0006         0
## rainannualpct -0.0250 0.0065   1e-04 0.0066  -0.0381   -0.0118         0
##
## $p_iv
```

```
## [1] 2
##
## $N
## [1] 1258733
##
## $N_cl
## NULL
##
## $df
## [1] 1258730
##
## $nvalues
##      represscount dissentcount lograin rainannualpct
## [1,]            3            5  390194        593785
```

```
plot_coef(g)
```



AJPS

Barth et al. (2015)

Replication Summary	
Unit of analysis	country*year

Replication Summary	
Treatment	wage inequality
Instrument	adjusted bargaining coverage; effective number of union confederations
Outcome	welfare support
Model	Table4(1)

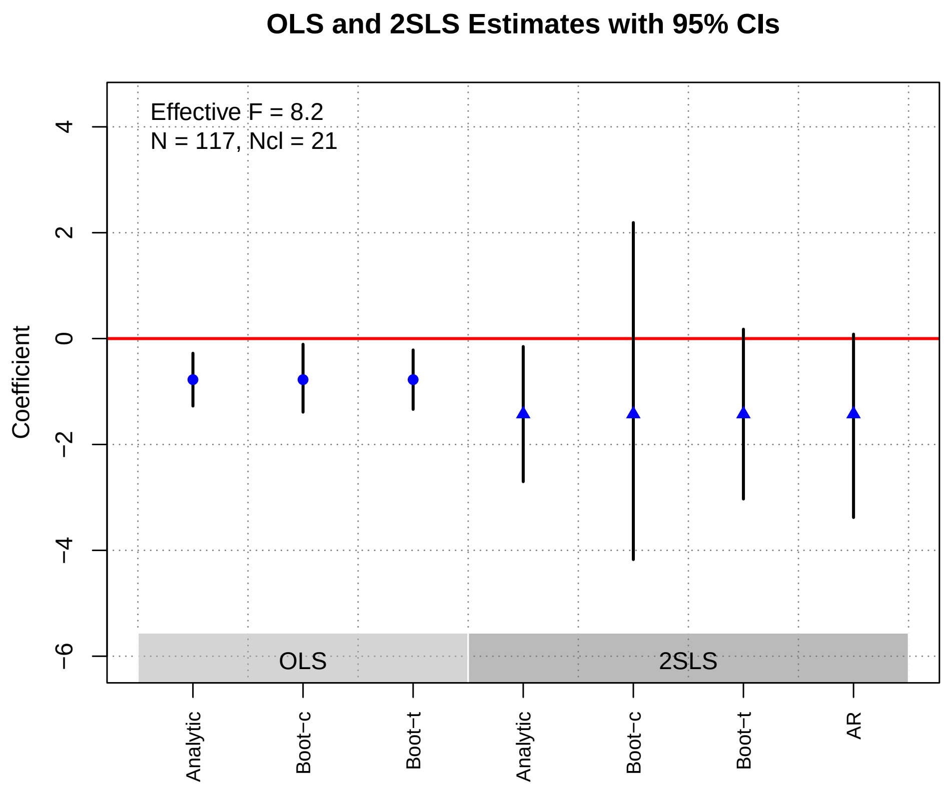
```
df<- readRDS("./data/ajps_Barth_2015.rds")
D <-"ld9d1"
Y <- "welfareleft"
Z <- c("l2ip_adjcov5", "l2ip_enucfs")
controls <- c("lgdpgr", "lelderly", "llntexp", "lud", "ludsq",
              "lechp", "lnet", "lannual", "ltrend", "ltrendsq")
cl <- FE <- "countrynumber"
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 30.794 sec.
## AR Test Inversion...
```

```
## $est_ols
##             Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -0.7755 0.2543 -3.0495 -1.2739  -0.2771  0.0023
## Boot.c   -0.7755 0.3232 -2.3995 -1.3894  -0.1073  0.0360
## Boot.t   -0.7755 0.2543 -3.0495 -1.3362  -0.2148  0.0050
##
## $est_2sls
##             Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -1.4265 0.6510 -2.1913 -2.7024  -0.1506  0.0284
## Boot.c   -1.4265 2.0891 -0.6828 -4.1740   2.1913  0.3160
## Boot.t   -1.4265 0.6510 -2.1913 -3.0302   0.1773  0.0810
##
## $AR
## $AR$Fstat
##        F      df1      df2        p
##   2.7758   2.0000 114.0000   0.0665
##
## $AR$ci.print
## [1] "[-3.3794, 0.0838]"
##
## $AR$ci
## [1] -3.37935086  0.08375199
##
```

```
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##      9.7741     15.0268     11.5754      3.2015      8.1611
##
## $rho
## [1] 0.4345
##
## $est_rf
##                 Coef    SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## l2ip_adjcov5 0.0184 0.013  0.1563 0.0192  -0.0258    0.0493     0.338
## l2ip_enucfs  0.1687 0.192  0.3797 0.4111  -0.8570    0.8513     0.714
##
## $est_fs
##                  Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## l2ip_adjcov5 -0.0096 0.0042  0.0228 0.0064  -0.0258   -0.0003     0.038
## l2ip_enucfs  -0.1542 0.0564  0.0063 0.1073  -0.2903    0.0716     0.182
##
## $p_iv
## [1] 2
##
## $N
## [1] 117
##
## $N_cl
## [1] 21
##
## $df
## [1] 20
##
## $nvalues
##      welfareleft ld9d1 l2ip_adjcov5 l2ip_enucfs
## [1,]         117   117          106         112
```

```
plot_coef(g)
```



Blair et al. (2022)

Replication Summary	
Unit of analysis	UN peacekeeping operations event level
Treatment	fragmentation of any given PKO mandate
Instrument	average fragmentation of all ongoing PKO mandates
Outcome	process performance
Model	TableD7(3)

```r
df <-readRDS("./data/ajps_Blair_2022.rds")
df<-as.data.frame(df)
D<-"L_avg"
Y <- "sh_perfassist_pb"
Z <- "L_fract_assistv3"
  controls <- c("L_experman_assist_pbv3","L_numtask_assist_pbv3","L_lntot",
                "L_deployment","L_lnpop","L_lngdp","L_ucdpconflictspell","L_polity")
cl <- NULL
FE <- c("date3","iso3n")
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

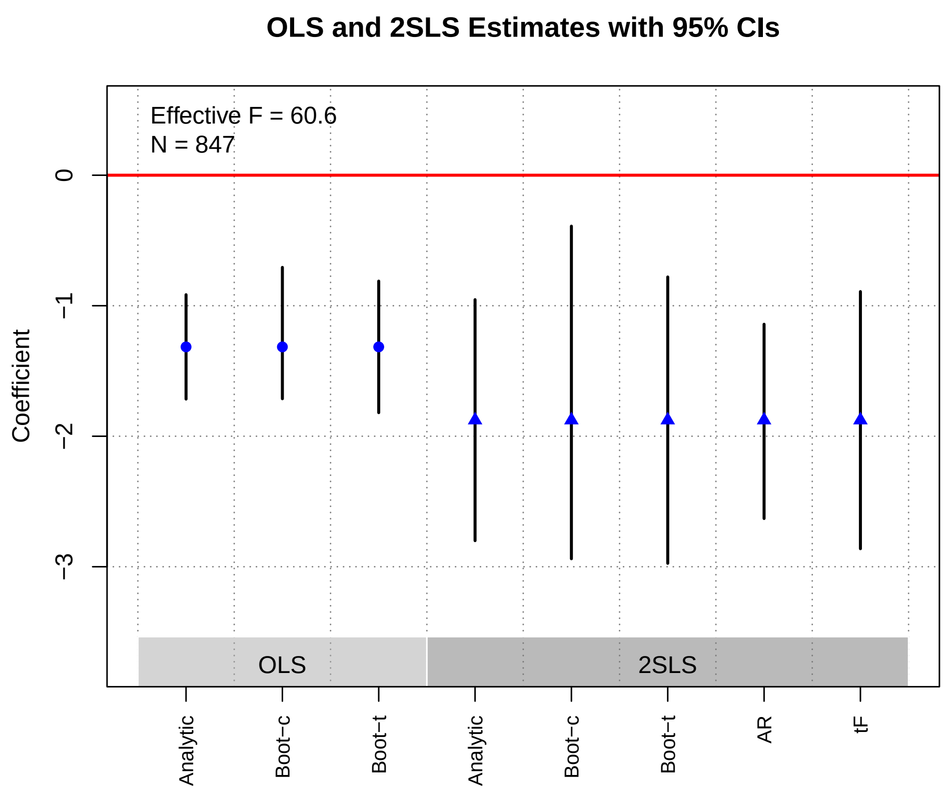
```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
```

```
## Bootstrap took 29.813 sec.
## AR Test Inversion...

## $est_ols
##             Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -1.3155 0.2040 -6.4481 -1.7153  -0.9156       0
## Boot.c   -1.3155 0.2568 -5.1219 -1.7128  -0.7063       0
## Boot.t   -1.3155 0.2040 -6.4481 -1.8195  -0.8115       0
##
## $est_2sls
##             Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -1.8768 0.4711 -3.9841 -2.8001  -0.9535  0.0001
## Boot.c   -1.8768 0.6478 -2.8970 -2.9389  -0.3904  0.0160
## Boot.t   -1.8768 0.4711 -3.9841 -2.9739  -0.7797  0.0020
##
## $AR
## $AR$Fstat
##        F      df1      df2        p
##  23.9745   1.0000 845.0000   0.0000
##
## $AR$ci.print
## [1] "[-2.6305, -1.1419]"
##
## $AR$ci
## [1] -2.630499 -1.141930
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##    186.0679     60.6442          NA     23.0898     60.6442
##
## $rho
## [1] 0.4793
##
## $tF
##       F      cF    Coef      SE       t  CI2.5% CI97.5% p-value
## 60.6442  2.0913 -1.8768  0.4711 -3.9841 -2.8619 -0.8917  0.0002
##
## $est_rf
##                   Coef    SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## L_fract_assistv3 1.805 0.464   1e-04 0.7319   0.3633    3.2495     0.016
##
## $est_fs
##                     Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
```

```
## L_fract_assistv3 -0.9617 0.1235       0 0.2001   -1.498   -0.7283         0
##
## $p_iv
## [1] 1
##
## $N
## [1] 847
##
## $N_cl
## NULL
##
## $df
## [1] 624
##
## $nvalues
##      sh_perfassist_pb L_avg L_fract_assistv3
## [1,]               56    55              222
```

```
plot_coef(g)
```



Carnegie and Marinov (2017)

Replication Summary	
Unit of analysis	country*year
Treatment	foreign aid
Instrument	being a former colony of one of the Council members
Outcome	CIRI Human Empowerment index
Model	Table1(2)

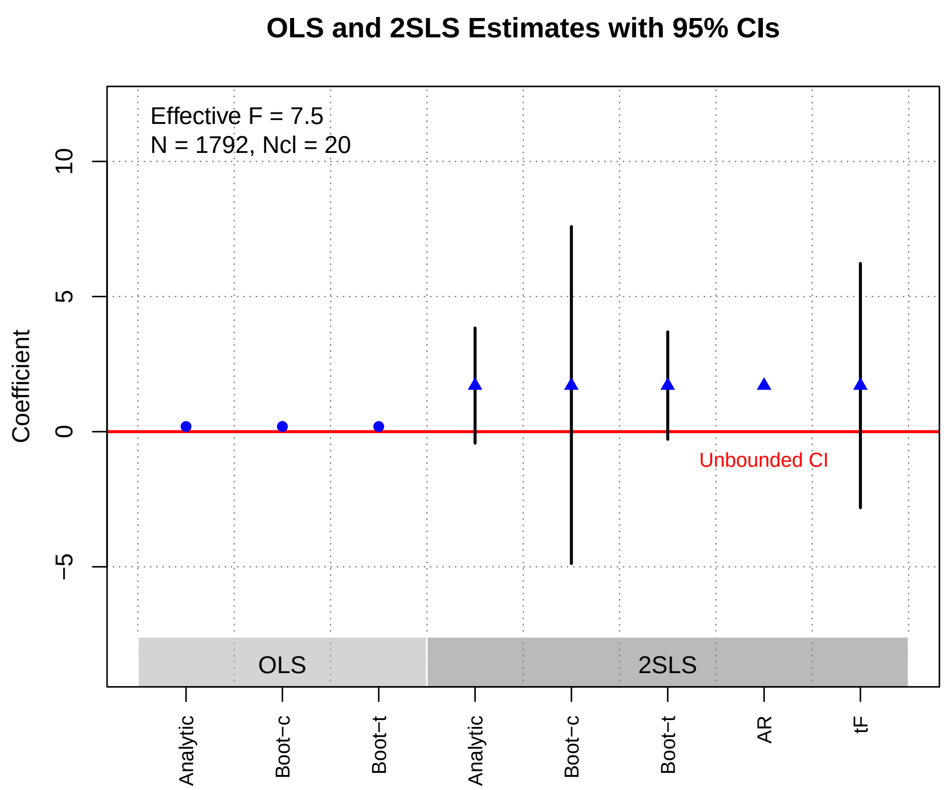
```
df<-readRDS("./data/ajps_Carnegie_etal_2017.rds")
D <-"EV"
Y <- "new_empinxavg"
Z <- "l2CPcol2"
controls <- c( "covloggdp", "covloggdpCF", "covloggdpC",
              "covdemregionF", "covdemregion", "coviNY_GDP_PETR_RT_ZSF",
              "coviNY_GDP_PETR_RT_ZS", "covwvs_relF", "covwvs_rel",
              "covwdi_imp", "covwdi_fdiF", "covwdi_fdi",
              "covwdi_expF", "covwdi_exp", "covihme_ayemF", "covihme_ayem")
cl<-c("year","ccode")
FE <- c("year","ccode")
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 31.897 sec.
## AR Test Inversion...

## $est_ols
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.1903 0.0578 3.2943  0.0771   0.3036   0.001
## Boot.c   0.1903 0.0751 2.5332  0.0531   0.3451   0.006
## Boot.t   0.1903 0.0578 3.2943  0.0423   0.3383   0.018
##
## $est_2sls
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 1.7054 1.0874 1.5684 -0.4259   3.8368  0.1168
## Boot.c   1.7054 4.4426 0.3839 -4.8803   7.5883  0.2080
## Boot.t   1.7054 1.0874 1.5684 -0.2823   3.6932  0.0760
##
## $AR
## $AR$Fstat
##         F       df1       df2         p
##    4.8424    1.0000 1790.0000    0.0279
##
## $AR$ci.print
## [1] "[0.2048, Inf)"
```

```
##
## $AR$ci
## [1] 0.204835      Inf
##
## $AR$bounded
## [1] FALSE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##      4.5101      4.5766      7.5007      3.7667      7.5007
##
## $rho
## [1] 0.0523
##
## $tF
##       F      cF    Coef      SE       t  CI2.5% CI97.5% p-value
##  7.5007  4.1570  1.7054  1.0874  1.5684 -2.8149  6.2258  0.4596
##
## $est_rf
##            Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## l2CPcol2 0.2632 0.1387  0.0578 0.1834  -0.0501    0.6223      0.12
##
## $est_fs
##            Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## l2CPcol2 0.1543 0.0721  0.0324 0.0795  -0.0388    0.2763     0.092
##
## $p_iv
## [1] 1
##
## $N
## [1] 1792
##
## $N_cl
## [1] 20
##
## $df
## [1] 19
##
## $nvalues
##      new_empinxavg   EV l2CPcol2
## [1,]            57 1601        2
```

```
plot_coef(g)
```



Chong et al. (2019)

Replication Summary	
Unit of analysis	household
Treatment	actual proportion of households treated in the locality
Instrument	treatment assignment in get-out-to-vote campaigns
Outcome	voted in 2013 presidential election
Model	Table4(1)

```
df <-readRDS("./data/ajps_Chong_etal_2019.rds")
D <-"ratio_treat"
Y <- "elecc_presid2013"
Z <- c("D2D30", "D2D40", "D2D50")
controls <-c("age", "married", "children", "num_children",
             "employed", "languag", "yrseduc", "bornloc",
             "hh_asset_index", "log_pop", "mujeres_perc",
             "pob_0_14_perc", "pob_15_64_perc", "pob_65mas_perc",
             "analfabetos_perc", "asiste_escuela_perc",
             "TASA_women", "TASA_men", "electricidad_perc",
             "agua_perc", "desague_perc", "basura_perc",
             "fono_fijo_perc", "fono_cel_perc", "ocupantes",
```

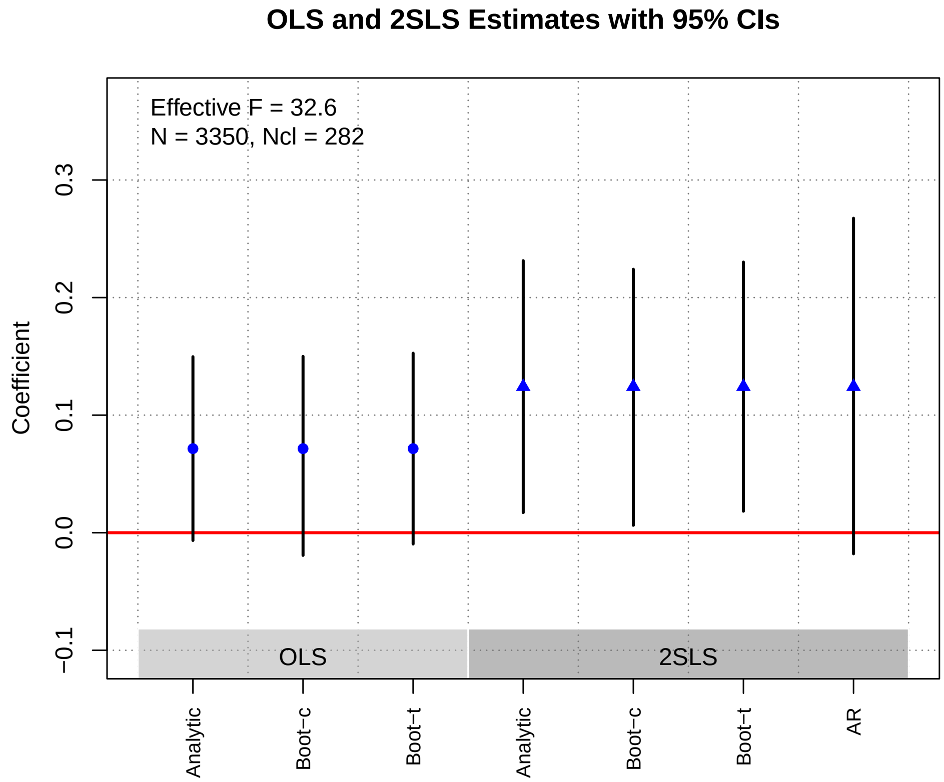
```r
             "Rural",  "distancia2_final", "db_age",
             "db_married", "db_children", "db_num_children",
             "db_employed", "db_languag", "db_yrseduc",
             "db_bornloc", "db_hh_asset_index", "db_log_pop",
             "db_mujeres_perc", "db_pob_0_14_perc",
             "db_pob_15_64_perc", "db_pob_65mas_perc",
             "db_analfabetos_perc", "db_asiste_escuela_perc",
             "db_TASA_women", "db_TASA_men", "db_electricidad_perc",
             "db_agua_perc", "db_desague_perc", "db_basura_perc",
             "db_fono_fijo_perc", "db_fono_cel_perc",
             "db_ocupantes", "db_Rural", "db_distancia2_final",
             "dpto1", "elecc_presid2008", "db_elecc_presid2008")
cl <- "loc"
FE <- NULL
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 25.124 sec.
## AR Test Inversion...

## $est_ols
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.0715 0.0399 1.7944 -0.0066   0.1497  0.0728
## Boot.c   0.0715 0.0442 1.6179 -0.0193   0.1500  0.1240
## Boot.t   0.0715 0.0399 1.7944 -0.0096   0.1527  0.0870
##
## $est_2sls
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.1242 0.0547 2.2719  0.0171   0.2314  0.0231
## Boot.c   0.1242 0.0548 2.2682  0.0063   0.2241  0.0380
## Boot.t   0.1242 0.0547 2.2719  0.0183   0.2302  0.0170
##
## $AR
## $AR$Fstat
##         F       df1       df2         p
##    2.0479    3.0000 3346.0000    0.1050
##
## $AR$ci.print
## [1] "[-0.0179, 0.2675]"
##
## $AR$ci
## [1] -0.01793968  0.26748144
##
## $AR$bounded
```

```
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##   1163.8658    270.5690     37.7653     34.8911     32.5611
##
## $rho
## [1] 0.7163
##
## $est_rf
##         Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## D2D30 0.0194 0.0321  0.5459 0.0353  -0.0539    0.0882     0.582
## D2D40 0.0651 0.0268  0.0150 0.0263   0.0097    0.1164     0.020
## D2D50 0.0190 0.0286  0.5075 0.0291  -0.0404    0.0714     0.514
##
## $est_fs
##         Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## D2D30 0.2996 0.0187       0 0.0470   0.2198    0.4030         0
## D2D40 0.3946 0.0220       0 0.0771   0.2531    0.5613         0
## D2D50 0.2663 0.0174       0 0.0452   0.1929    0.3652         0
##
## $p_iv
## [1] 3
##
## $N
## [1] 3350
##
## $N_cl
## [1] 282
##
## $df
## [1] 3316
##
## $nvalues
##      elecc_presid2013 ratio_treat D2D30 D2D40 D2D50
## [1,]                2          56     2     2     2
```

```
plot_coef(g)
```



Colantone and Stanig (2018)

Replication Summary	
Unit of analysis	region*year
Treatment	regional import shock from China
Instrument	Chinese imports to the United States
Outcome	Economic nationalism
Model	Table1(1)

```
df <-readRDS("./data/ajps_Colantone_etal_2018.rds")
D <-"import_shock"
Y <- "median_nationalism"
Z <- "instrument_for_shock"
controls <- NULL
cl <- "nuts2_year"
FE <- "fix_effect"
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

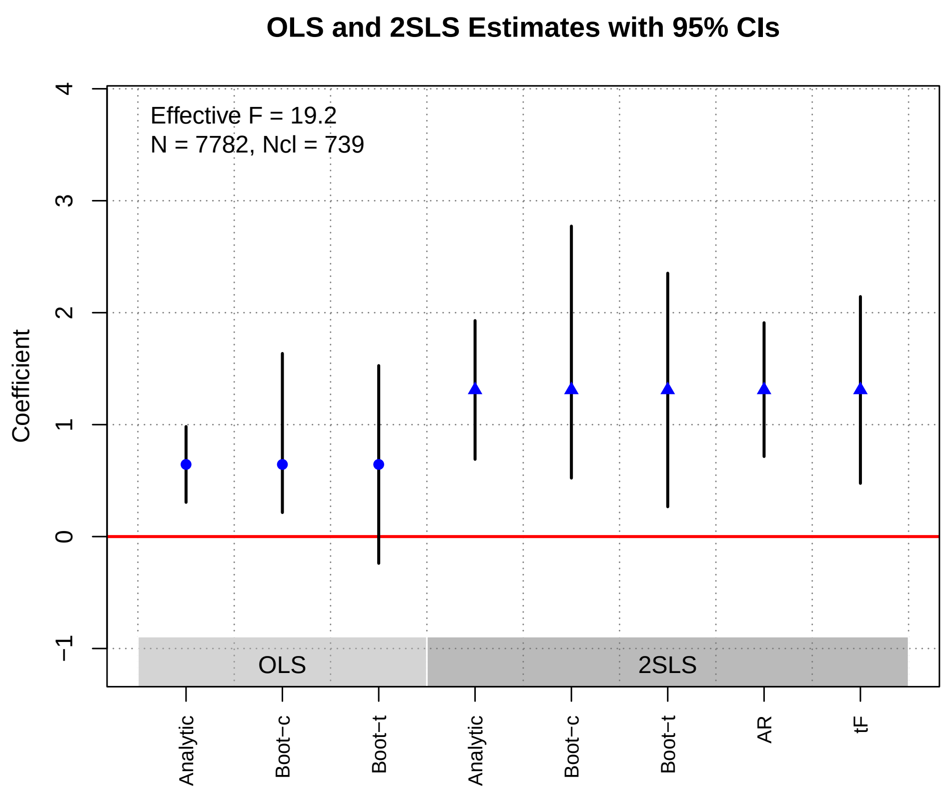
```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 32.869 sec.
## AR Test Inversion...
```

```
## Parallelising on 15 cores

## $est_ols
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.6442 0.1726 3.7326  0.3059   0.9825  0.0002
## Boot.c   0.6442 0.3679 1.7509  0.2150   1.6353  0.0000
## Boot.t   0.6442 0.1726 3.7326 -0.2388   1.5272  0.1100
##
## $est_2sls
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 1.3096 0.3160 4.1437  0.6901   1.9290   0.000
## Boot.c   1.3096 0.5694 2.3000  0.5228   2.7730   0.000
## Boot.t   1.3096 0.3160 4.1437  0.2666   2.3525   0.023
##
## $AR
## $AR$Fstat
##         F       df1       df2         p
##   18.3415    1.0000 7780.0000    0.0000
##
## $AR$ci.print
## [1] "[0.7154, 1.9100]"
##
## $AR$ci
## [1] 0.715422 1.910047
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##   1810.3678     42.8350     19.1709     11.8485     19.1709
##
## $rho
## [1] 0.4358
##
## $tF
##       F      cF    Coef      SE       t  CI2.5% CI97.5% p-value
## 19.1709  2.6386  1.3096  0.3160  4.1437  0.4757  2.1435  0.0021
##
## $est_rf
##                        Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## instrument_for_shock 0.0514 0.0093       0 0.0202   0.0238    0.1025         0
##
## $est_fs
##                        Coef    SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## instrument_for_shock 0.0392 0.006       0 0.0114   0.0258    0.0704         0
```

```
##
## $p_iv
## [1] 1
##
## $N
## [1] 7782
##
## $N_cl
## [1] 739
##
## $df
## [1] 7724
##
## $nvalues
##      median_nationalism import_shock instrument_for_shock
## [1,]                167          739                  739
```

```
plot_coef(g)
```



Coppock and Green (2016)

Replication Summary	
Unit of analysis	individual
Treatment	voting in November 2007 municipal elections

Replication Summary	
Instrument	mailing showing 2005 Vote
Outcome	voting in the 2008 presidential primary
Model	Table2(2)

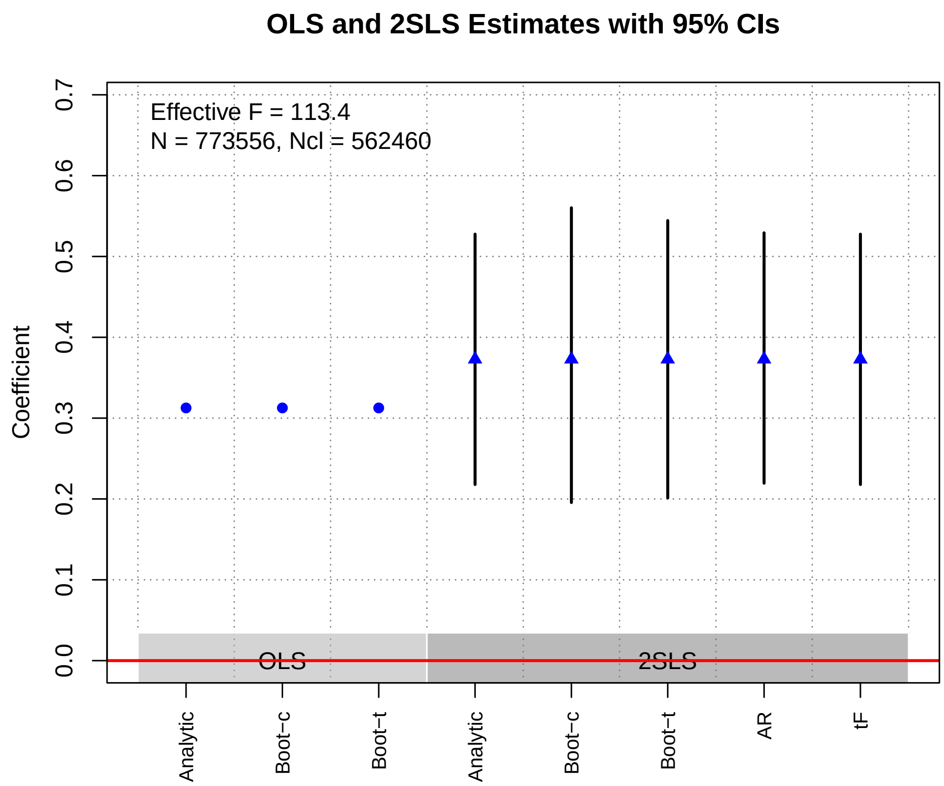
```
df<-readRDS("./data/ajps_Coppock_etal_2016.rds")
D <-"og2007"
Y <- "JAN2008"
Z <- "treat2"
controls <- NULL
cl <- "hh"
FE <- NULL
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 19.119 sec.
## AR Test Inversion...
## Parallelising on 15 cores
```

```
## $est_ols
##             Coef     SE        t CI 2.5% CI 97.5% p.value
## Analytic 0.3126 0.0012 258.2569  0.3102   0.3149       0
## Boot.c   0.3126 0.0013 234.4950  0.3099   0.3151       0
## Boot.t   0.3126 0.0012 258.2569  0.3100   0.3152       0
##
## $est_2sls
##             Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.3728 0.0790 4.7183  0.2179   0.5276       0
## Boot.c   0.3728 0.0926 4.0257  0.1956   0.5602       0
## Boot.t   0.3728 0.0790 4.7183  0.2012   0.5444       0
##
## $AR
## $AR$Fstat
##           F         df1         df2           p
##     20.8095      1.0000 773554.0000      0.0000
##
## $AR$ci.print
## [1] "[0.2195, 0.5292]"
##
## $AR$ci
## [1] 0.2195093 0.5292273
##
## $AR$bounded
```

```
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##    165.8659    151.8337    113.3680    113.1763    113.3680
##
## $rho
## [1] 0.0146
##
## $tF
##        F       cF     Coef       SE        t   CI2.5%  CI97.5%  p-value
## 113.3680   1.9600   0.3728   0.0790   4.7183   0.2179   0.5276   0.0000
##
## $est_rf
##          Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## treat2 0.0187 0.0041       0 0.0047   0.0094    0.0282         0
##
## $est_fs
##          Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## treat2 0.0502 0.0041       0 0.0047   0.0406    0.0597         0
##
## $p_iv
## [1] 1
##
## $N
## [1] 773556
##
## $N_cl
## [1] 562460
##
## $df
## [1] 773554
##
## $nvalues
##      JAN2008 og2007 treat2
## [1,]       2      2      2
```

```
plot_coef(g)
```



De La O (2013)

Replication Summary	
Unit of analysis	village
Treatment	early coverage of Conditional Cash Transfer
Instrument	random assignment to early coverage
Outcome	incumbent party's vote share
Model	Table3(b1)

```
df <- readRDS("./data/ajps_De_La_O_2013.rds")
D <-"early_progresa_p"
Y <- "t2000"
Z <- "treatment"
controls <- c("avgpoverty","pobtot1994", "votos_totales1994",
              "pri1994", "pan1994", "prd1994")
cl <- NULL
FE <- "villages"
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

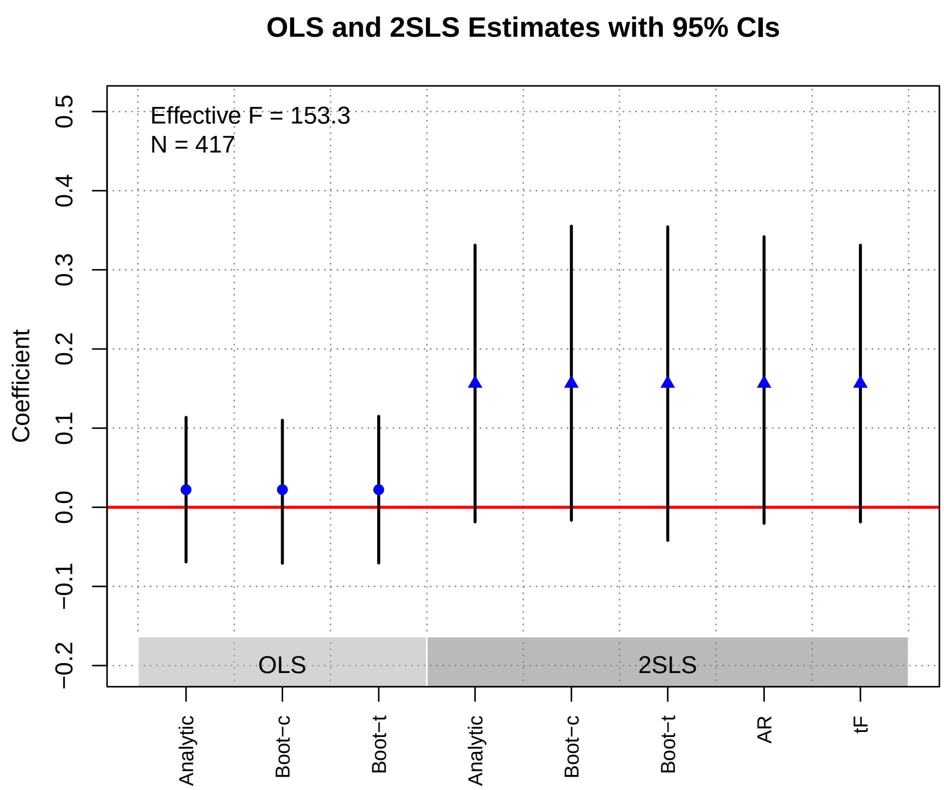
```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 31.618 sec.
```

```
## AR Test Inversion...

## $est_ols
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.0222 0.0466 0.4771 -0.0691   0.1136  0.6333
## Boot.c   0.0222 0.0459 0.4846 -0.0708   0.1100  0.7240
## Boot.t   0.0222 0.0466 0.4771 -0.0704   0.1149  0.6430
##
## $est_2sls
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.1563 0.0892 1.7521 -0.0185   0.3312  0.0798
## Boot.c   0.1563 0.0942 1.6584 -0.0164   0.3552  0.0800
## Boot.t   0.1563 0.0892 1.7521 -0.0418   0.3544  0.1110
##
## $AR
## $AR$Fstat
##        F      df1      df2        p
##   2.9752   1.0000 415.0000   0.0853
##
## $AR$ci.print
## [1] "[-0.0203, 0.3419]"
##
## $AR$ci
## [1] -0.02033352  0.34185831
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##    177.1916    153.2854          NA    147.3725    153.2854
##
## $rho
## [1] 0.556
##
## $tF
##        F       cF     Coef       SE        t   CI2.5%  CI97.5%  p-value
## 153.2854   1.9600   0.1563   0.0892   1.7521  -0.0185   0.3312   0.0798
##
## $est_rf
##             Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## treatment 0.0532 0.0296  0.0723 0.0306  -0.0058    0.1141      0.08
##
## $est_fs
##             Coef     SE p.value  SE.b CI.b2.5% CI.b97.5% p.value.b
## treatment 0.3401 0.0275       0 0.028   0.2869    0.3964         0
```

```
##
## $p_iv
## [1] 1
##
## $N
## [1] 417
##
## $N_cl
## NULL
##
## $df
## [1] 396
##
## $nvalues
##      t2000 early_progresa_p treatment
## [1,]   407              251         2
```

```
plot_coef(g)
```



Goldstein and You (2017)

Replication Summary	
Unit of analysis	city
Treatment	lobbying spending

Replication Summary	
Instrument	direct flight to Washington, DC
Outcome	total earmarks or grants awarded
Model	Table4(4)

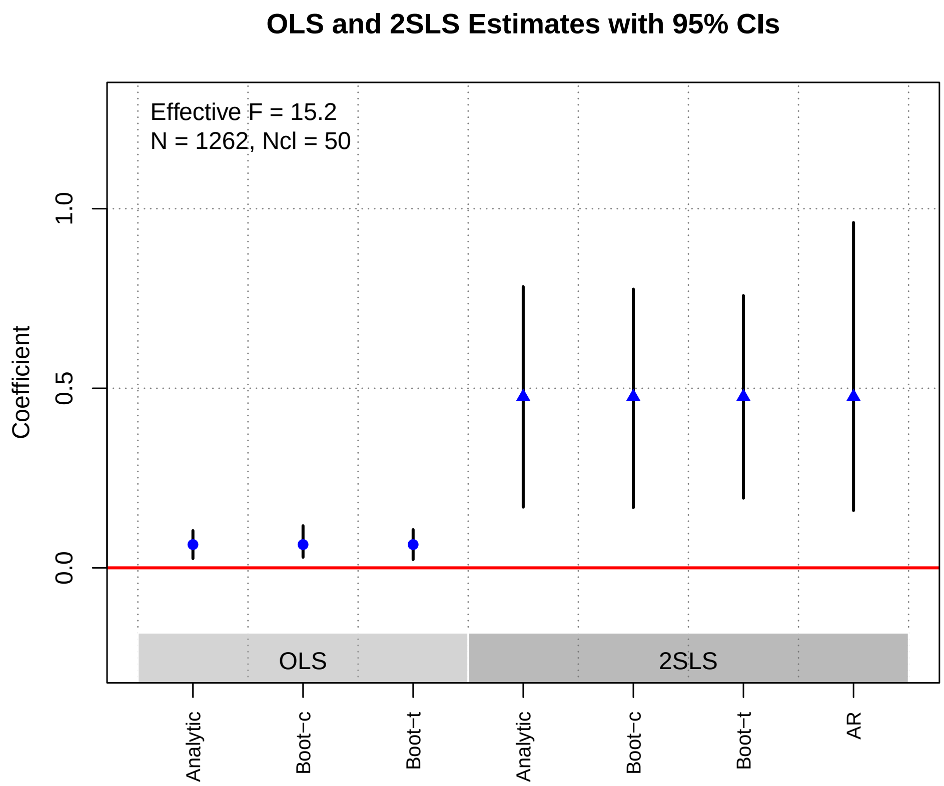
```
df <- readRDS("./data/ajps_Goldstein_etal_2017.rds")
df <- as.data.frame(df)
Y <-"ln_recovery"
D <-"ln_citylob"
Z <- c("direct_flight_dc", "diverge2_r")
controls <- c("pop_r", "land_r", "water_r", "senior_r", "student_r", "ethnic_r",
              "mincome_r", "unemp_r", "poverty_r", "gini_r", "city_propertytaxshare_r",
              "city_intgovrevenueshare_r", "city_airexp_r", "houdem_r", "ln_countylob")
cl <- "state2"
FE <- "state2"
weights <- NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl, weights=weights, cores = cores, parallel = TRUE))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 30.749 sec.
## AR Test Inversion...

## $est_ols
##             Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.0648 0.0198 3.2692  0.0259   0.1036  0.0011
## Boot.c   0.0648 0.0223 2.9075  0.0296   0.1173  0.0000
## Boot.t   0.0648 0.0198 3.2692  0.0231   0.1064  0.0030
##
## $est_2sls
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.476 0.1566 3.0407  0.1692   0.7829  0.0024
## Boot.c   0.476 0.1550 3.0709  0.1680   0.7762  0.0060
## Boot.t   0.476 0.1566 3.0407  0.1942   0.7579  0.0020
##
## $AR
## $AR$Fstat
##         F       df1       df2         p
##    6.8022    2.0000 1259.0000    0.0012
##
## $AR$ci.print
## [1] "[0.1598, 0.9614]"
##
## $AR$ci
## [1] 0.1597997 0.9613764
```

```
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##     16.6195     13.7688     15.7426     14.4539     15.1587
##
## $rho
## [1] 0.1645
##
## $est_rf
##                    Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## direct_flight_dc 1.2403 0.5540  0.0252 0.6194  -0.2766    2.1686     0.102
## diverge2_r       0.3010 0.1588  0.0581 0.1783  -0.0277    0.6606     0.074
##
## $est_fs
##                    Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## direct_flight_dc 2.6658 0.6544  0.0000 0.7607   0.9735    4.0004     0.004
## diverge2_r       0.6070 0.2065  0.0033 0.2405   0.1563    1.1154     0.004
##
## $p_iv
## [1] 2
##
## $N
## [1] 1262
##
## $N_cl
## [1] 50
##
## $df
## [1] 49
##
## $nvalues
##      ln_recovery ln_citylob direct_flight_dc diverge2_r
## [1,]        1196        135                2       1262
```

```
plot_coef(g)
```



Hager and Hilbig (2019) a

Replication Summary	
Unit of analysis	city
Treatment	equiTable inheritance customs
Instrument	mean elevation
Outcome	female representation
Model	Table3(1)

```
df<-readRDS("./data/ajps_Hager_etal_2019.rds")
D <-"fair_dic"
Y <- "gem_women_share"
Z <- "elev_mean"
controls <- c("lon", "lat", "childlabor_mean_1898",
              "support_expenses_total_capita","gem_council",
              "gem_pop_density","pop_tot")
cl<- NULL
FE<- c("state2","law_cat2")
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

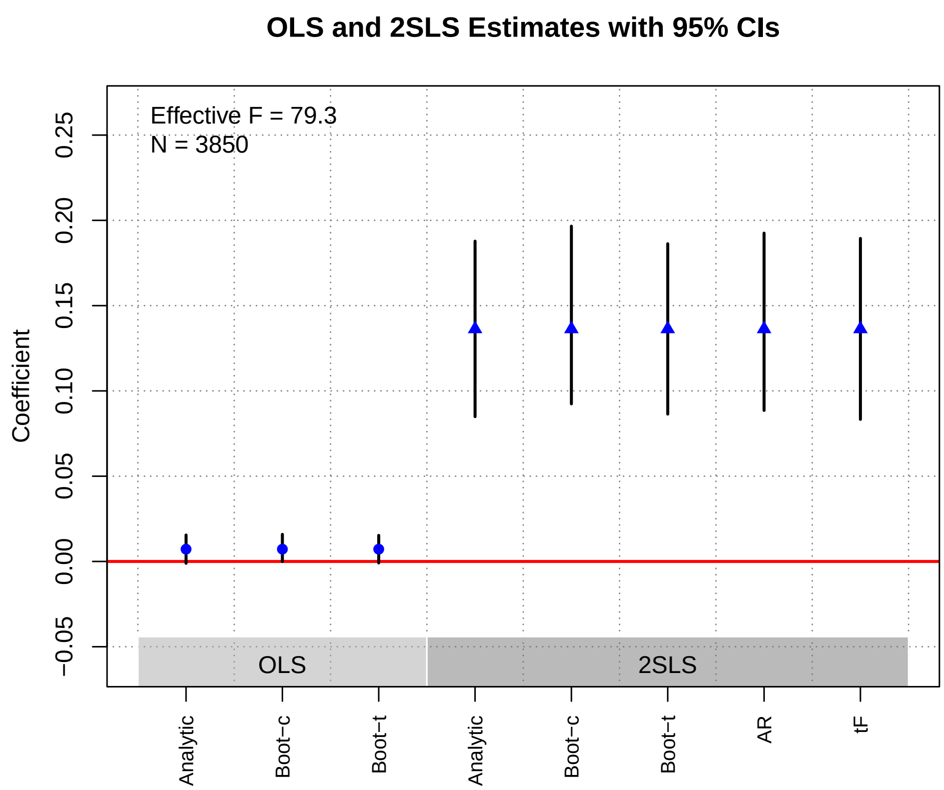
```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
```

```
## Bootstrap took 30.675 sec.
## AR Test Inversion...

## $est_ols
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.0072 0.0042 1.7010 -0.0011   0.0155  0.0889
## Boot.c   0.0072 0.0042 1.7192 -0.0001   0.0159  0.0580
## Boot.t   0.0072 0.0042 1.7010 -0.0009   0.0153  0.0810
##
## $est_2sls
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.1363 0.0262 5.1939  0.0849   0.1878       0
## Boot.c   0.1363 0.0270 5.0547  0.0924   0.1966       0
## Boot.t   0.1363 0.0262 5.1939  0.0864   0.1863       0
##
## $AR
## $AR$Fstat
##        F      df1      df2        p
##   34.311    1.000 3848.000    0.000
##
## $AR$ci.print
## [1] "[0.0886, 0.1925]"
##
## $AR$ci
## [1] 0.08856418 0.19251398
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##    122.1930     79.2985          NA     83.6605     79.2985
##
## $rho
## [1] 0.1758
##
## $tF
##       F      cF    Coef      SE       t  CI2.5% CI97.5% p-value
## 79.2985  2.0200  0.1363  0.0262  5.1939  0.0833  0.1894  0.0000
##
## $est_rf
##             Coef SE p.value SE.b CI.b2.5% CI.b97.5% p.value.b
## elev_mean -1e-04  0       0    0   -2e-04    -1e-04         0
##
## $est_fs
##             Coef    SE p.value  SE.b CI.b2.5% CI.b97.5% p.value.b
```

```
## elev_mean -9e-04 1e-04       0 1e-04  -0.0012    -7e-04         0
##
## $p_iv
## [1] 1
##
## $N
## [1] 3850
##
## $N_cl
## NULL
##
## $df
## [1] 3831
##
## $nvalues
##      gem_women_share fair_dic elev_mean
## [1,]             230        2      3850
```

```
plot_coef(g)
```



Hager and Hilbig (2019) b

Replication Summary	
Unit of analysis	city
Treatment	equiTable inheritance customs
Instrument	distance to rivers
Outcome	female representation
Model	Table3(2)

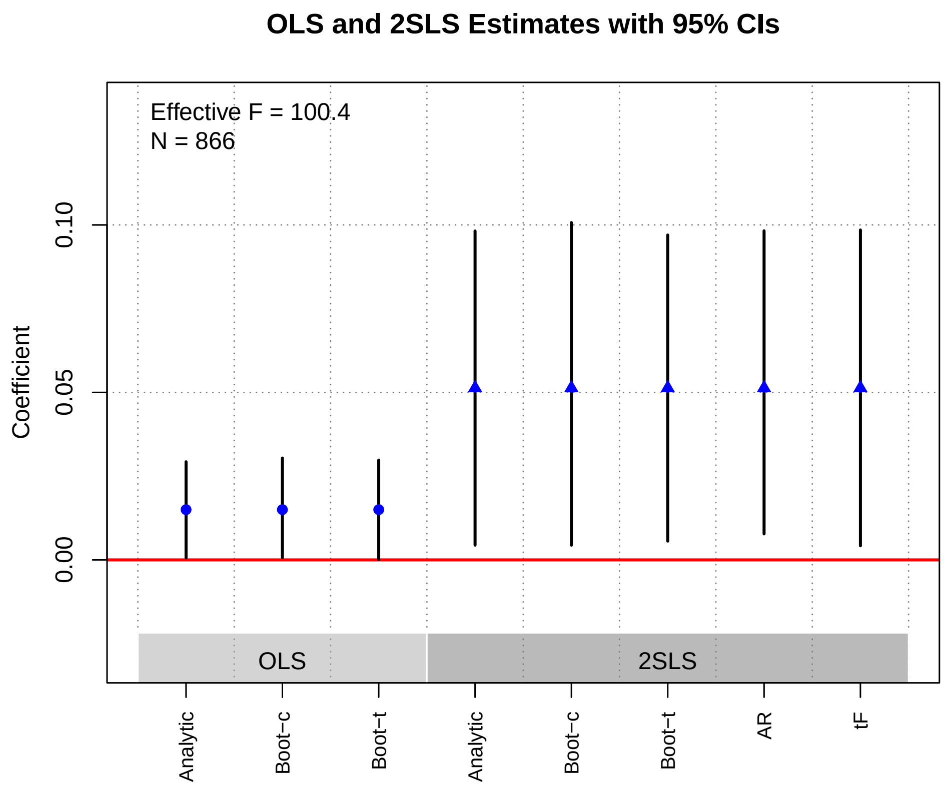
```
df<-readRDS("./data/ajps_Hager_etal_2019.rds")
D <-"fair_dic"
Y <- "gem_women_share"
Z <-"river_dist_min"
controls <- c("lon", "lat", "childlabor_mean_1898",
              "support_expenses_total_capita","gem_council",
              "gem_pop_density","pop_tot")
cl<- NULL
FE<- c("law_cat2")
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 31.772 sec.
## AR Test Inversion...
```

```
## $est_ols
##           Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.015 0.0073 2.0379   6e-04   0.0293  0.0416
## Boot.c   0.015 0.0074 2.0159   7e-04   0.0304  0.0400
## Boot.t   0.015 0.0073 2.0379   1e-04   0.0298  0.0480
##
## $est_2sls
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.0513 0.0239 2.1441  0.0044   0.0982   0.032
## Boot.c   0.0513 0.0249 2.0576  0.0044   0.1007   0.028
## Boot.t   0.0513 0.0239 2.1441  0.0056   0.0970   0.027
##
## $AR
## $AR$Fstat
##        F      df1      df2        p
##   5.2375   1.0000 864.0000   0.0223
##
## $AR$ci.print
## [1] "[0.0078, 0.0982]"
##
## $AR$ci
```

```
## [1] 0.007758233 0.098236346
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##     99.1676    100.3609          NA     88.5516    100.3609
##
## $rho
## [1] 0.3222
##
## $tF
##        F       cF     Coef       SE        t   CI2.5%  CI97.5%  p-value
## 100.3609   1.9700   0.0513   0.0239   2.1441   0.0042   0.0985   0.0329
##
## $est_rf
##                   Coef    SE p.value  SE.b CI.b2.5% CI.b97.5% p.value.b
## river_dist_min -5e-04 2e-04  0.0291 3e-04   -0.001         0     0.028
##
## $est_fs
##                    Coef    SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## river_dist_min -0.0105 0.001       0 0.0011  -0.0126   -0.0084         0
##
## $p_iv
## [1] 1
##
## $N
## [1] 866
##
## $N_cl
## NULL
##
## $df
## [1] 856
##
## $nvalues
##      gem_women_share fair_dic river_dist_min
## [1,]             110        2            866
```

```
plot_coef(g)
```



Hong et al. (2022)

Replication Summary	
Unit of analysis	township
Treatment	NVM subsidy per voter
Instrument	Terrain elevation slope
Outcome	Park's vote share in 2012
Model	Table3(3)

```
df <-readRDS("./data/ajps_Hong_etal_2022.rds")
df<-as.data.frame(df)
D<-"total_Lamount_1974_1978_perelect"
Y <- "E18ConsSh"
Z <- c("te_median1", "ts_median1")
controls <- c("area_1970","demo_female_share_1966","demo_age_15plus_1966",
              "demo_illiterate_1966","demo_pop_ch_1970_1966","E17ConsSh","eup")
cl <- "CTY_cd"
FE <- "CTY_cd"
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

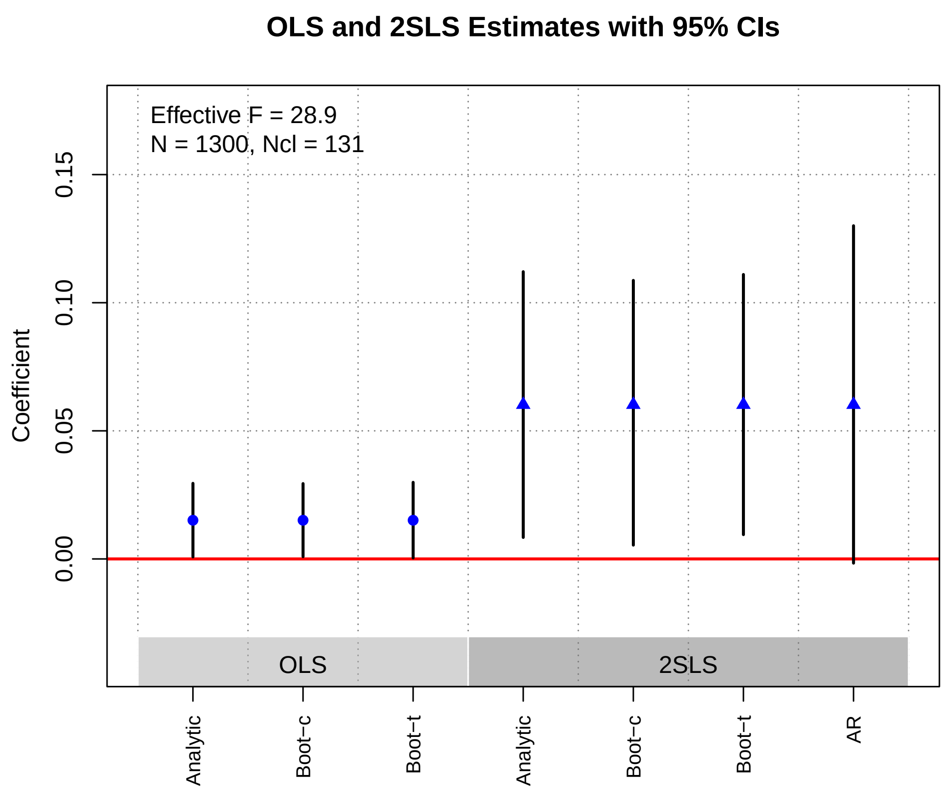
```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
```

```
## Bootstrap took 31.714 sec.
## AR Test Inversion...

## $est_ols
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.0151 0.0073 2.0652   8e-04   0.0295  0.0389
## Boot.c   0.0151 0.0073 2.0849   9e-04   0.0294  0.0320
## Boot.t   0.0151 0.0073 2.0652   4e-04   0.0299  0.0460
##
## $est_2sls
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.0602 0.0264 2.2781  0.0084   0.1121  0.0227
## Boot.c   0.0602 0.0262 2.2966  0.0054   0.1087  0.0220
## Boot.t   0.0602 0.0264 2.2781  0.0095   0.1110  0.0240
##
## $AR
## $AR$Fstat
##         F       df1       df2         p
##    2.8694    2.0000 1297.0000    0.0571
##
## $AR$ci.print
## [1] "[-0.0016, 0.1300]"
##
## $AR$ci
## [1] -0.00163529  0.13002933
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##     34.7064     29.0832     28.2296     28.1521     28.8604
##
## $rho
## [1] 0.2376
##
## $est_rf
##               Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## te_median1 -0.0036 0.0232  0.8771 0.0231  -0.0526    0.0404     0.826
## ts_median1  0.0020 0.0011  0.0622 0.0010   0.0001    0.0040     0.038
##
## $est_fs
##              Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## te_median1 0.3276 0.1262  0.0094 0.1402   0.0763    0.6116     0.016
## ts_median1 0.0171 0.0054  0.0017 0.0064   0.0053    0.0301     0.004
##
```

```
## $p_iv
## [1] 2
##
## $N
## [1] 1300
##
## $N_cl
## [1] 131
##
## $df
## [1] 130
##
## $nvalues
##      E18ConsSh total_Lamount_1974_1978_perelect te_median1 ts_median1
## [1,]      1292                             1285       1300       1232
```

```
plot_coef(g)
```



Kim (2019)

Replication Summary	
Unit of analysis	municipality*year
Treatment	Democratic institutions
Instrument	population threshold

Replication Summary	
Outcome	women political engagement
Model	Table2(1)

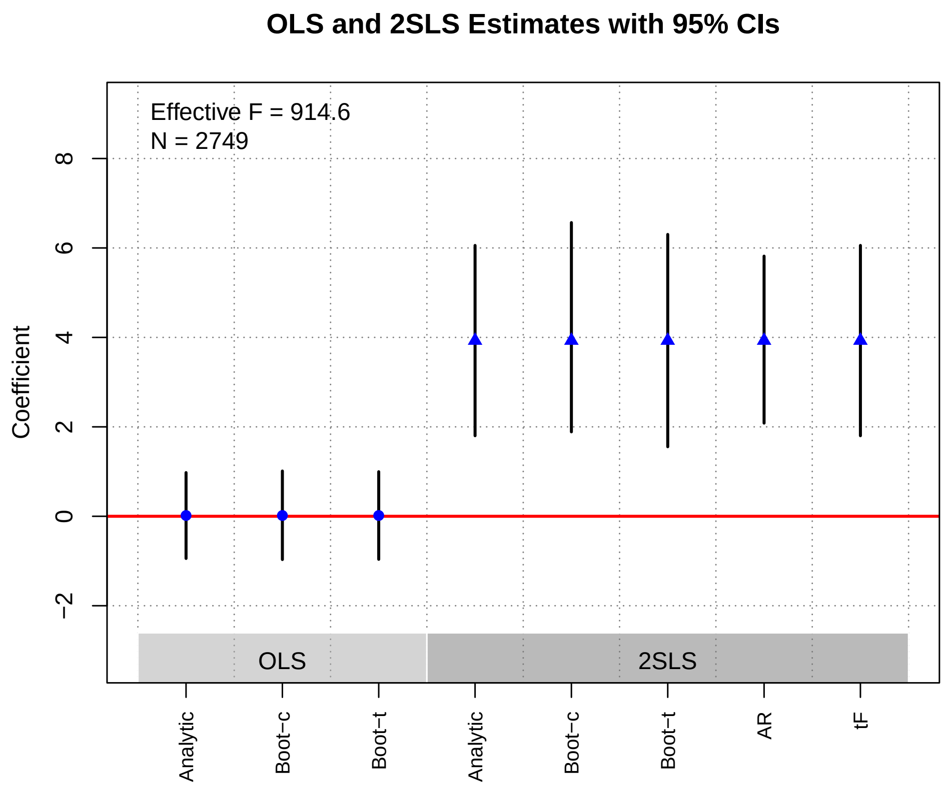
```
df<- readRDS("./data/ajps_Kim_2019.rds")
D <-"direct"
Y <- "wm_turnout"
Z <-  "new"
controls <- c("left", "wm_voters", "enep")
cl <- NULL
FE <- "year"
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 31.616 sec.
## AR Test Inversion...
```

```
## $est_ols
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.017 0.4897 0.0346 -0.9429   0.9768  0.9724
## Boot.c   0.017 0.5016 0.0338 -0.9673   1.0124  0.9420
## Boot.t   0.017 0.4897 0.0346 -0.9625   0.9965  0.9680
##
## $est_2sls
##             Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 3.9287 1.0855 3.6192  1.8011   6.0563   3e-04
## Boot.c   3.9287 1.1882 3.3063  1.8880   6.5672   0e+00
## Boot.t   3.9287 1.0855 3.6192  1.5548   6.3025   1e-03
##
## $AR
## $AR$Fstat
##         F       df1       df2         p
##   17.3741    1.0000 2747.0000    0.0000
##
## $AR$ci.print
## [1] "[2.0833, 5.8175]"
##
## $AR$ci
## [1] 2.083289 5.817457
##
## $AR$bounded
## [1] TRUE
##
```

```
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##   1007.3382    914.6461          NA    857.6002    914.6461
##
## $rho
## [1] 0.5186
##
## $tF
##        F       cF     Coef       SE        t   CI2.5%  CI97.5%  p-value
## 914.6461   1.9600   3.9287   1.0855   3.6192   1.8011   6.0563   0.0003
##
## $est_rf
##      Coef    SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## new 1.949 0.516   2e-04 0.5534   0.9688    3.1216         0
##
## $est_fs
##       Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## new 0.4961 0.0164       0 0.0169    0.458    0.5256         0
##
## $p_iv
## [1] 1
##
## $N
## [1] 2749
##
## $N_cl
## NULL
##
## $df
## [1] 2738
##
## $nvalues
##      wm_turnout direct new
## [1,]       2606      2   2
```

```
plot_coef(g)
```



Kocher et al. (2011)

Replication Summary	
Unit of analysis	hamlet (smallest population unit)
Treatment	aerial bombing
Instrument	past insurgent control
Outcome	changes in local control
Model	Table5(5B)

```
df<-readRDS("./data/ajps_Kocher_etal_2011.rds")
D <-"bombed_969"
Y<- "mod2a_1adec"
Z <- c("mod2a_1ajul", "mod2a_1aaug")
controls <- c("mod2a_1asep", "score", "ln_dist", "std", "lnhpop")
cl<- NULL
FE <-NULL
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

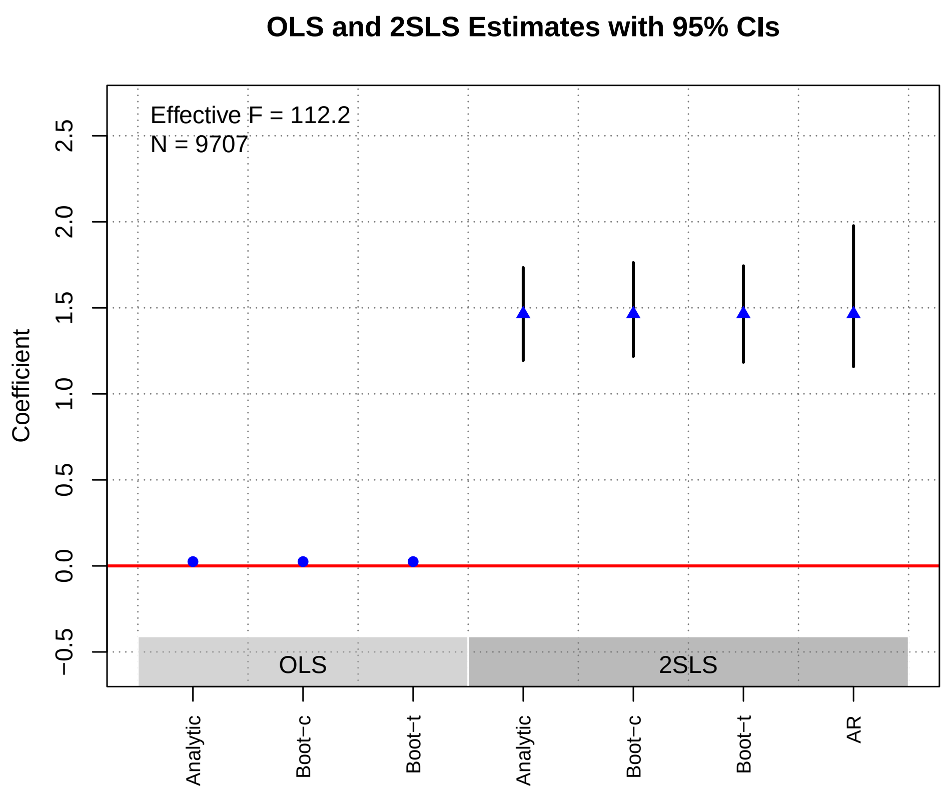
```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 38.176 sec.
## AR Test Inversion...
```

```
## Parallelising on 15 cores

## $est_ols
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.0249 0.0042 5.8926  0.0166   0.0332       0
## Boot.c   0.0249 0.0042 5.8618  0.0179   0.0341       0
## Boot.t   0.0249 0.0042 5.8926  0.0163   0.0335       0
##
## $est_2sls
##           Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic 1.464 0.1377 10.6345  1.1942   1.7339       0
## Boot.c   1.464 0.1415 10.3463  1.2181   1.7629       0
## Boot.t   1.464 0.1377 10.6345  1.1839   1.7441       0
##
## $AR
## $AR$Fstat
##         F       df1       df2         p
##  936.1377    2.0000 9704.0000    0.0000
##
## $AR$ci.print
## [1] "[1.1584, 1.9771]"
##
## $AR$ci
## [1] 1.158401 1.977146
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##     44.1703     59.8861          NA     59.6058    112.1923
##
## $rho
## [1] 0.095
##
## $est_rf
##               Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## mod2a_1ajul 0.2562 0.0123       0 0.0118   0.2342    0.2786         0
## mod2a_1aaug 0.1830 0.0134       0 0.0132   0.1570    0.2085         0
##
## $est_fs
##               Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## mod2a_1ajul 0.1681 0.0284       0 0.0272   0.1139    0.2227         0
## mod2a_1aaug 0.1328 0.0311       0 0.0309   0.0721    0.1963         0
##
## $p_iv
```

```
## [1] 2
##
## $N
## [1] 9707
##
## $N_cl
## NULL
##
## $df
## [1] 9700
##
## $nvalues
##      mod2a_1adec bombed_969 mod2a_1ajul mod2a_1aaug
## [1,]           5         35           5           5
```

```
plot_coef(g)
```



Lelkes et al. (2017)

Replication Summary	
Unit of analysis	state*year
Treatment	number of broadband Internet providers
Instrument	state-level ROW index
Outcome	affective polarization

Replication Summary	
Model	Table1(3)

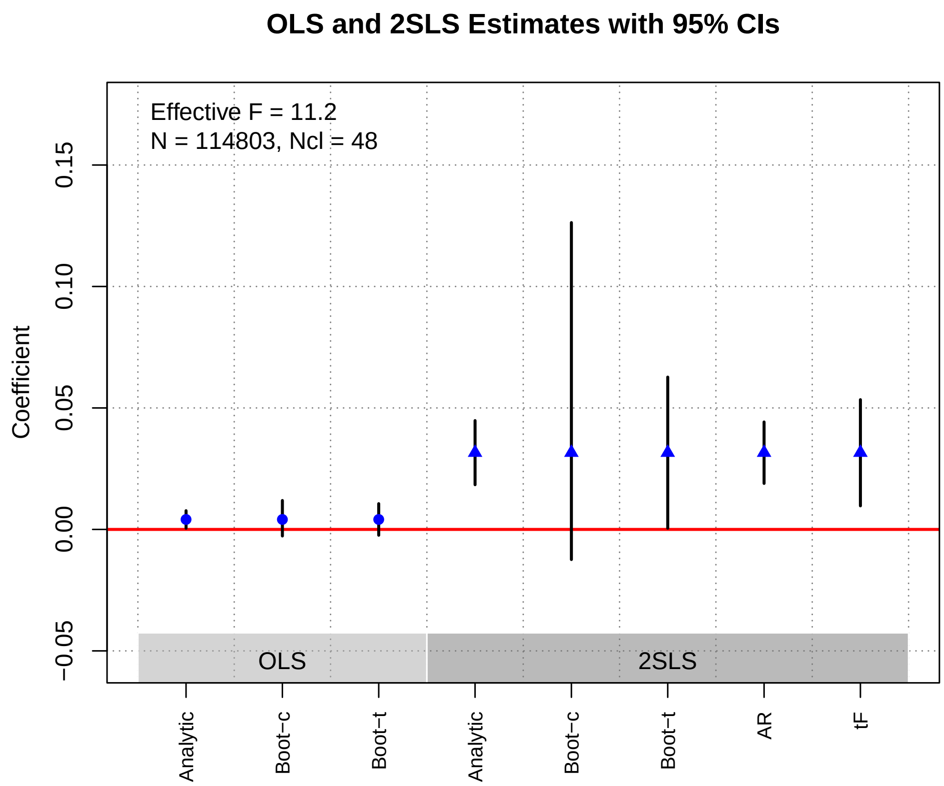
```
df<-readRDS("./data/ajps_Lelkes_2017.rds")
D <-"D"
Y <- "outcome"
Z <- "Total_log"
controls <- c("region", "percent_black", "percent_white",
              "percent_male", "lowed", "unemploymentrate",
              "density", "HHINC_log")
cl<- "state"
FE <- "year"
weights=NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 1.662 sec.
## AR Test Inversion...
## Parallelising on 15 cores

## $est_ols
##             Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.0041 0.0018 2.2577  0.0005   0.0077   0.024
## Boot.c   0.0041 0.0036 1.1410 -0.0027   0.0119   0.238
## Boot.t   0.0041 0.0018 2.2577 -0.0024   0.0106   0.205
##
## $est_2sls
##             Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.0316 0.0067 4.6941  0.0184   0.0448   0.000
## Boot.c   0.0316 0.1193 0.2647 -0.0124   0.1263   0.088
## Boot.t   0.0316 0.0067 4.6941  0.0005   0.0627   0.047
##
## $AR
## $AR$Fstat
##           F         df1         df2           p
##     23.6566      1.0000 114801.0000      0.0000
##
## $AR$ci.print
## [1] "[0.0189, 0.0442]"
##
## $AR$ci
## [1] 0.01893067 0.04422426
##
## $AR$bounded
```

```
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##   9525.8467   8161.7346     11.1632      7.4289     11.1632
##
## $rho
## [1] 0.2768
##
## $tF
##       F      cF    Coef      SE       t  CI2.5% CI97.5% p-value
## 11.1632  3.2489  0.0316  0.0067  4.6941  0.0097  0.0534  0.0046
##
## $est_rf
##             Coef    SE p.value  SE.b CI.b2.5% CI.b97.5% p.value.b
## Total_log 0.0033 7e-04       0 0.002   -3e-04    0.0079     0.072
##
## $est_fs
##             Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## Total_log 0.1042 0.0012       0 0.0382   0.0171    0.1707      0.02
##
## $p_iv
## [1] 1
##
## $N
## [1] 114803
##
## $N_cl
## [1] 48
##
## $df
## [1] 114790
##
## $nvalues
##      outcome    D Total_log
## [1,]    2423 1438        43
```

```
plot_coef(g)
```



López-Moctezuma et al. (2020)

Replication Summary	
Unit of analysis	individual
Treatment	town-hall meetings
Instrument	assignment to treatment
Outcome	voting behavior
Model	figure3(2)

```
df <-readRDS("./data/ajps_Moctezuma_etal_2020.rds")
df<-as.data.frame(df)
D<-"treatment"
Y <- "vote"
Z <- "assignment"
  controls <- NULL
cl <- "barangay"
FE <- "city"
weights<-"weight.att"
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

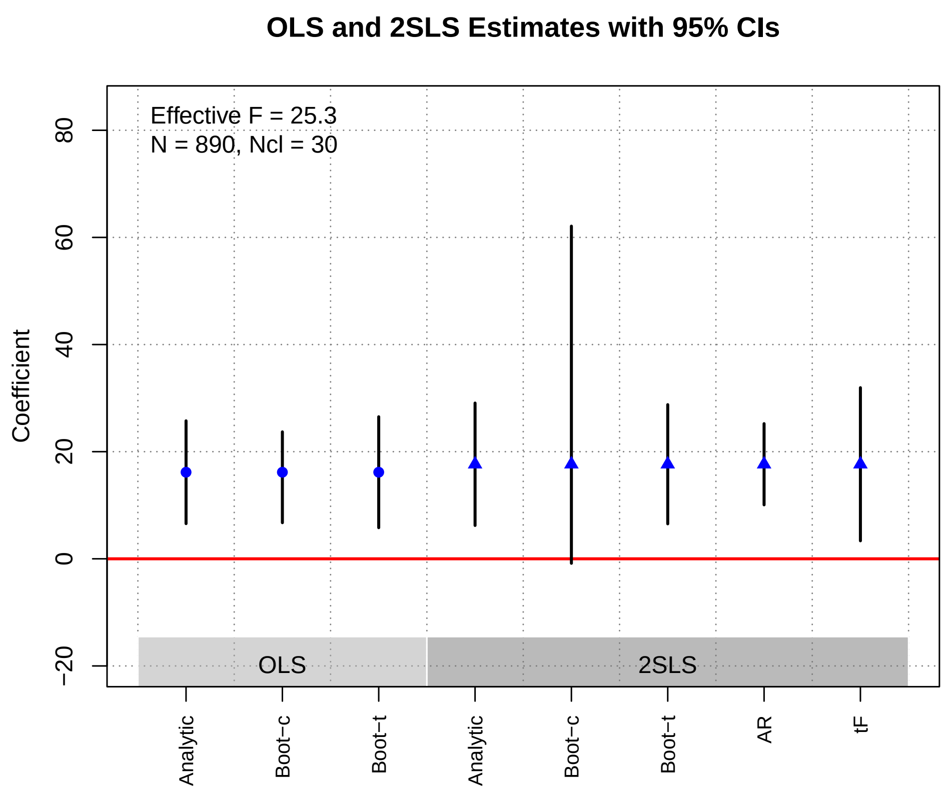
```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 31.646 sec.
```

```
## AR Test Inversion...

## $est_ols
##             Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 16.1643 4.8955 3.3019  6.5692  25.7594   0.001
## Boot.c   16.1643 4.3166 3.7447  6.7402  23.7014   0.002
## Boot.t   16.1643 4.8955 3.3019  5.8084  26.5201   0.024
##
## $est_2sls
##             Coef       SE      t CI 2.5% CI 97.5% p.value
## Analytic 17.6531   5.8296 3.0282  6.2271  29.0791  0.0025
## Boot.c   17.6531 233.4686 0.0756 -0.8380  62.1023  0.0521
## Boot.t   17.6531   5.8296 3.0282  6.5283  28.7779  0.0300
##
## $AR
## $AR$Fstat
##        F      df1      df2        p
##  20.3497   1.0000 888.0000   0.0000
##
## $AR$ci.print
## [1] "[10.0746, 25.2315]"
##
## $AR$ci
## [1] 10.07461 25.23153
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##   1663.9064    521.4034     25.2694      6.1571     25.2694
##
## $rho
## [1] 0.8089
##
## $tF
##       F      cF    Coef      SE       t  CI2.5% CI97.5% p-value
## 25.2694  2.4519 17.6531  5.8296  3.0282  3.3593 31.9469  0.0155
##
## $est_rf
##               Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## assignment 13.2179 4.3439  0.0023 6.3727    1.151   26.1623     0.028
##
## $est_fs
##              Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## assignment 0.7488 0.0328       0 0.3018  -0.0162         1    0.0601
```

```
##
## $p_iv
## [1] 1
##
## $N
## [1] 890
##
## $N_cl
## [1] 30
##
## $df
## [1] 879
##
## $nvalues
##      vote treatment assignment
## [1,]    2         2          2
```

```
plot_coef(g)
```



McClendon (2014)

Replication Summary	
Unit of analysis	individual
Treatment	reading social esteem promising email

Replication Summary	
Instrument	assignment to treatment
Outcome	participation in LGBTQ events
Model	Table2(1)

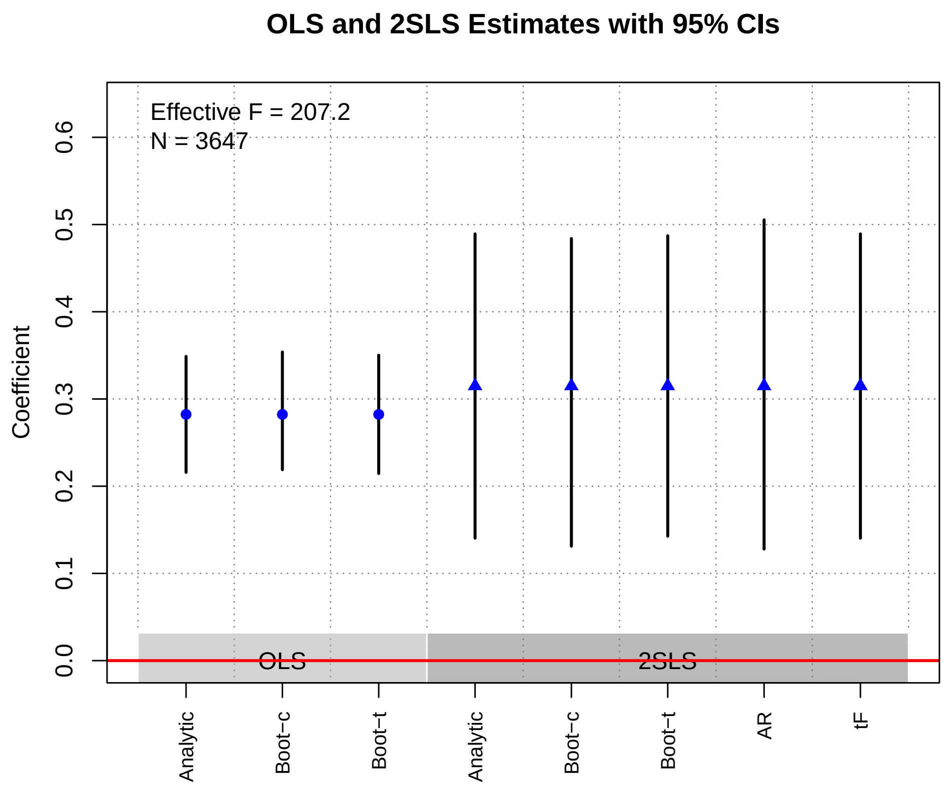
```
df <- readRDS("./data/ajps_McClendon_2014.rds")
D<-"openedesteem"
Y<- "intended"
Z <- "esteem"
controls <- NULL
cl<- NULL
FE <- NULL
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 12.683 sec.
## AR Test Inversion...
```

```
## $est_ols
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.2823 0.0339 8.3291  0.2159   0.3488       0
## Boot.c   0.2823 0.0345 8.1866  0.2189   0.3539       0
## Boot.t   0.2823 0.0339 8.3291  0.2145   0.3502       0
##
## $est_2sls
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.3149 0.0890 3.5376  0.1404   0.4893   4e-04
## Boot.c   0.3149 0.0903 3.4855  0.1312   0.4840   0e+00
## Boot.t   0.3149 0.0890 3.5376  0.1427   0.4871   0e+00
##
## $AR
## $AR$Fstat
##         F       df1       df2         p
##   10.1309    1.0000 3645.0000    0.0015
##
## $AR$ci.print
## [1] "[0.1280, 0.5054]"
##
## $AR$ci
## [1] 0.1279570 0.5053555
##
## $AR$bounded
## [1] TRUE
```

```
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##    103.7604    207.1798          NA    211.5796    207.1798
##
## $rho
## [1] 0.1664
##
## $tF
##        F       cF     Coef       SE        t   CI2.5%  CI97.5%  p-value
## 207.1798   1.9600   0.3149   0.0890   3.5376   0.1404   0.4893   0.0004
##
## $est_rf
##          Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## esteem 0.0247 0.0072   5e-04 0.0071   0.0103    0.0372         0
##
## $est_fs
##          Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## esteem 0.0786 0.0055       0 0.0054   0.0682    0.0892         0
##
## $p_iv
## [1] 1
##
## $N
## [1] 3647
##
## $N_cl
## NULL
##
## $df
## [1] 3645
##
## $nvalues
##      intended openedesteem esteem
## [1,]        2            2      2
```

```
plot_coef(g)
```



Rueda (2017)

Replication Summary	
Unit of analysis	city
Treatment	actual polling place size.
Instrument	the size of the polling station
Outcome	citizens' reports of electoral manipulation
Model	Table5(1)

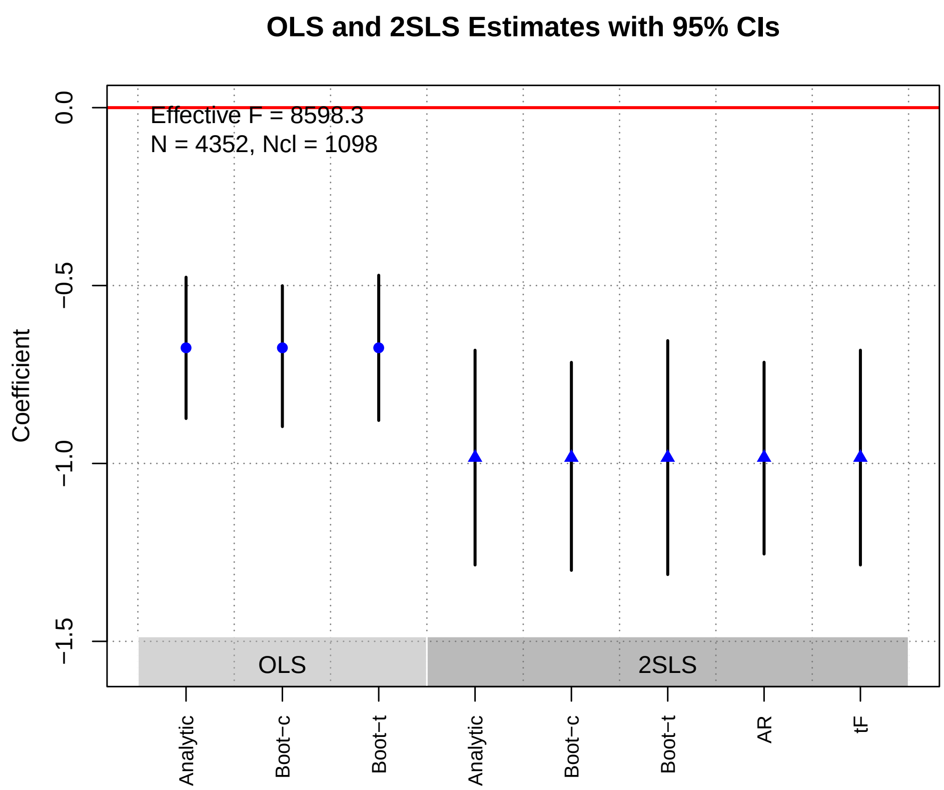
```
df <- readRDS("./data/ajps_Rueda_2017.rds")
D <-"lm_pob_mesa"
Y <- "e_vote_buying"
Z <- "lz_pob_mesa_f"
controls <- c("lpopulation", "lpotencial")
cl <- "muni_code"
FE <- NULL
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 14.063 sec.
## AR Test Inversion...
```

```
## $est_ols
##             Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -0.675 0.1013 -6.6630 -0.8736  -0.4765       0
## Boot.c   -0.675 0.0997 -6.7677 -0.8964  -0.5005       0
## Boot.t   -0.675 0.1013 -6.6630 -0.8791  -0.4710       0
##
## $est_2sls
##              Coef    SE       t CI 2.5% CI 97.5% p.value
## Analytic -0.9835 0.154 -6.3872 -1.2853  -0.6817       0
## Boot.c   -0.9835 0.146 -6.7348 -1.3004  -0.7159       0
## Boot.t   -0.9835 0.154 -6.3872 -1.3122  -0.6548       0
##
## $AR
## $AR$Fstat
##         F       df1       df2         p
##   50.5097    1.0000 4350.0000    0.0000
##
## $AR$ci.print
## [1] "[-1.2545, -0.7156]"
##
## $AR$ci
## [1] -1.2545169 -0.7155854
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##    3106.387    3108.591    8598.326    8700.525    8598.326
##
## $rho
## [1] 0.6455
##
## $tF
##         F        cF      Coef        SE         t    CI2.5%   CI97.5%   p-value
## 8598.3264    1.9600   -0.9835    0.1540   -6.3872   -1.2853   -0.6817    0.0000
##
## $est_rf
##                   Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## lz_pob_mesa_f -0.7826 0.1219       0 0.1153   -1.032   -0.5747         0
##
## $est_fs
##                 Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## lz_pob_mesa_f 0.7957 0.0143       0 0.0085   0.7801     0.813         0
##
## $p_iv
```

```
## [1] 1
##
## $N
## [1] 4352
##
## $N_cl
## [1] 1098
##
## $df
## [1] 4348
##
## $nvalues
##      e_vote_buying lm_pob_mesa lz_pob_mesa_f
## [1,]            16        4118          3860
```

```
plot_coef(g)
```



Sexton et al. (2019)

Replication Summary	
Unit of analysis	department*year
Treatment	health budget
Instrument	soldier fatalities
Outcome	health social service

Replication Summary	
Model	Table3(1)

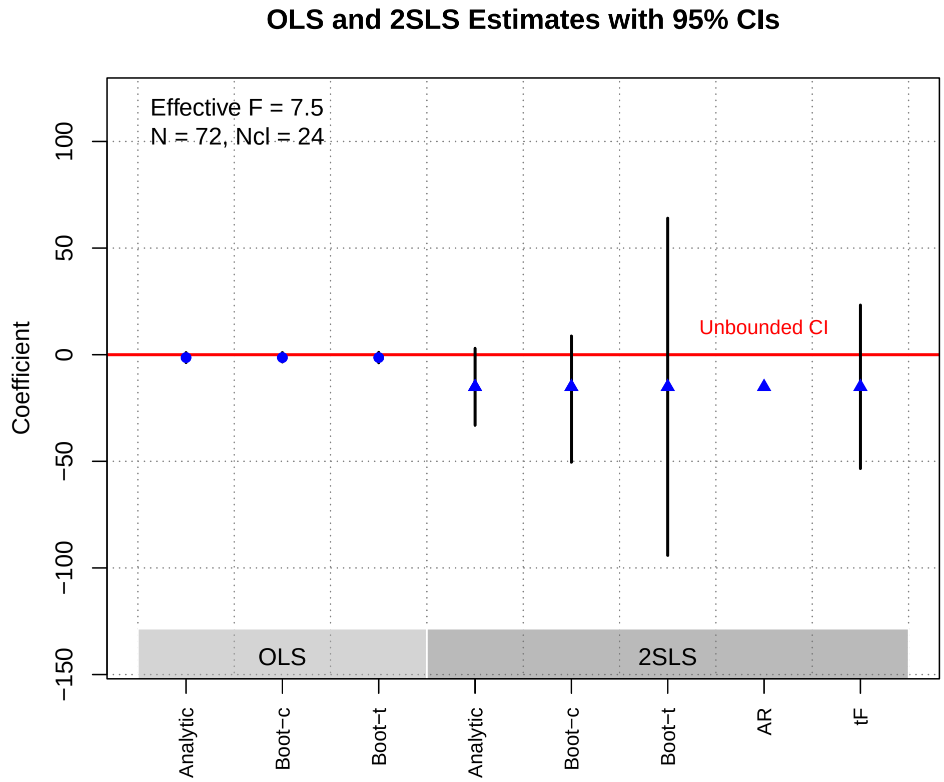
```
df <-readRDS("./data/ajps_Sexton_etal_2019.rds")
D<-"socialservice_b"
Y <- "Finfant_mortality"
Z <- "Lgk_budget"
controls <- c("Lgk_prebudget", "ln_pbi_pc", "execution_nohealth")
cl <- "deptcode"
FE <- c("year","deptcode")
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 30.628 sec.
## AR Test Inversion...
```

```
## $est_ols
##               Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -1.3472 1.1972 -1.1253 -3.6938   0.9994  0.2605
## Boot.c   -1.3472 1.1142 -1.2091 -3.4213   1.0000  0.2581
## Boot.t   -1.3472 1.1972 -1.1253 -3.7920   1.0975  0.2480
##
## $est_2sls
##                Coef      SE       t  CI 2.5% CI 97.5% p.value
## Analytic -15.0645  9.2117 -1.6354 -33.1195   2.9906  0.1020
## Boot.c   -15.0645 33.5214 -0.4494 -50.4262   8.7481  0.2114
## Boot.t   -15.0645  9.2117 -1.6354 -94.1248  63.9959  0.2348
##
## $AR
## $AR$Fstat
##       F     df1     df2       p
##  7.1494  1.0000 70.0000  0.0093
##
## $AR$ci.print
## [1] "(-Inf, -3.2734] Union [21.8384, Inf)"
##
## $AR$ci
## [1]      -Inf -3.273432 21.838356       Inf
##
## $AR$bounded
## [1] FALSE
##
##
```

```
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##      1.0172      2.5692      7.4923      2.7593      7.4923
##
## $rho
## [1] 0.1538
##
## $tF
##        F       cF     Coef       SE        t   CI2.5%  CI97.5%  p-value
##   7.4923   4.1607 -15.0645   9.2117  -1.6354 -53.3920  23.2630   0.4411
##
## $est_rf
##              Coef     SE p.value  SE.b CI.b2.5% CI.b97.5% p.value.b
## Lgk_budget 4.3552 1.5163  0.0041 2.137  -2.1134    6.1216    0.1911
##
## $est_fs
##               Coef     SE p.value  SE.b CI.b2.5% CI.b97.5% p.value.b
## Lgk_budget -0.2891 0.1804   0.109 0.174  -0.6643    0.0243    0.0549
##
## $p_iv
## [1] 1
##
## $N
## [1] 72
##
## $N_cl
## [1] 24
##
## $df
## [1] 23
##
## $nvalues
##      Finfant_mortality socialservice_b Lgk_budget
## [1,]                39              72          6
```

```
plot_coef(g)
```



Spenkuch and Tillmann (2018)

Replication Summary	
Unit of analysis	electoral district
Treatment	religion of voters living in the same areas more than three and a half centuries later
Instrument	individual princes' decisions concerning whether to adopt Protestantism
Outcome	Nazi vote share
Model	Table2(B1)

```
df <-readRDS("./data/ajps_Spenkuch_etal_2018.rds")
D <-"r_1925C_kath"
Y <- "r_NSDAP_NOV1932_p"
Z <- c("r_kath1624", "r_gem1624")
controls <- c("r_1925C_juden", "r_1925C_others",
              "r_M1925C_juden","r_M1925C_others")
cl <- 'WKNR'
FE <- NULL
weights="r_wahlberechtigte_NOV1932"
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

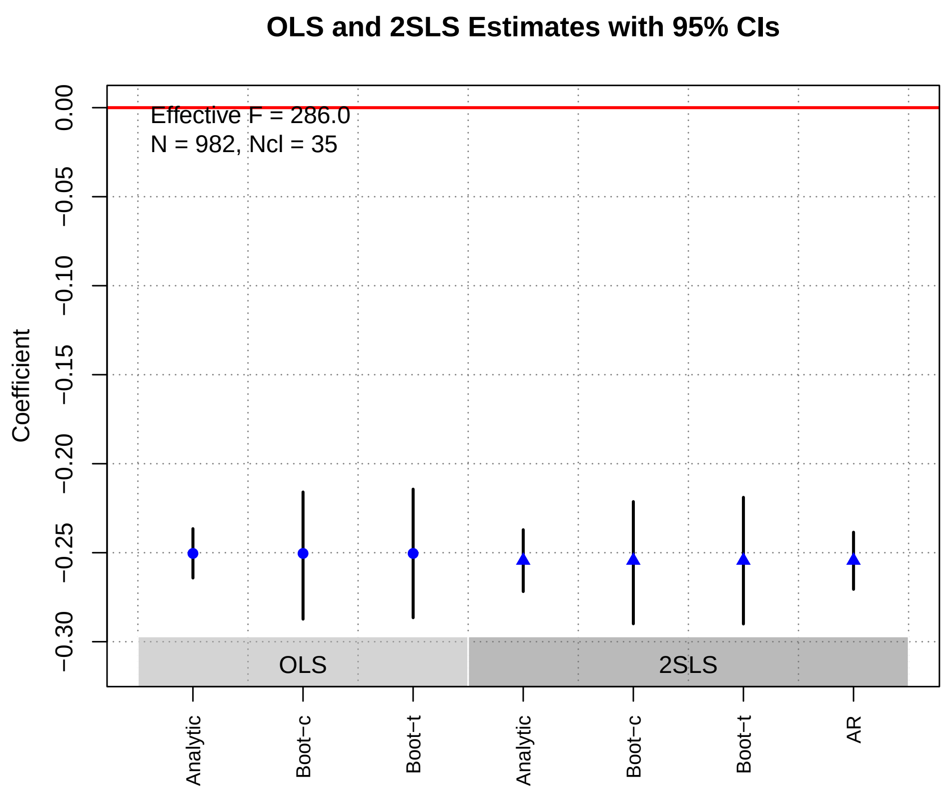
```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
```

```
## Bootstrap took 12.972 sec.
## AR Test Inversion...

## $est_ols
##             Coef     SE        t CI 2.5% CI 97.5% p.value
## Analytic -0.2504 0.0071 -35.4593 -0.2642  -0.2365       0
## Boot.c   -0.2504 0.0179 -13.9727 -0.2873  -0.2159       0
## Boot.t   -0.2504 0.0071 -35.4593 -0.2865  -0.2143       0
##
## $est_2sls
##             Coef     SE        t CI 2.5% CI 97.5% p.value
## Analytic -0.2544 0.0089 -28.7257 -0.2718  -0.2371       0
## Boot.c   -0.2544 0.0175 -14.5760 -0.2899  -0.2213       0
## Boot.t   -0.2544 0.0089 -28.7257 -0.2900  -0.2189       0
##
## $AR
## $AR$Fstat
##        F      df1      df2        p
## 361.9389   2.0000 979.0000   0.0000
##
## $AR$ci.print
## [1] "[-0.2706, -0.2385]"
##
## $AR$ci
## [1] -0.2705518 -0.2384884
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##   1215.3547    726.7058    212.7390    200.7118    286.0263
##
## $rho
## [1] 0.8446
##
## $est_rf
##                Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## r_kath1624 -17.2028 0.7546       0 1.2702 -19.6313  -14.6714         0
## r_gem1624   -9.1477 1.2001       0 1.6802 -12.8906   -6.2825         0
##
## $est_fs
##               Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## r_kath1624 66.6657 1.7581       0 3.3309  59.7794   73.1142         0
## r_gem1624  39.2697 3.1667       0 4.8034  31.2867   49.7198         0
##
```

```
## $p_iv
## [1] 2
##
## $N
## [1] 982
##
## $N_cl
## [1] 35
##
## $df
## [1] 978
##
## $nvalues
##      r_NSDAP_NOV1932_p r_1925C_kath r_kath1624 r_gem1624
## [1,]               982          977          2         2
```

```
plot_coef(g)
```



Stokes (2016)

Replication Summary	
Unit of analysis	precinct
Treatment	turbine location
Instrument	wind speed

Replication Summary	
Outcome	vote turnout
Model	Table2(2)

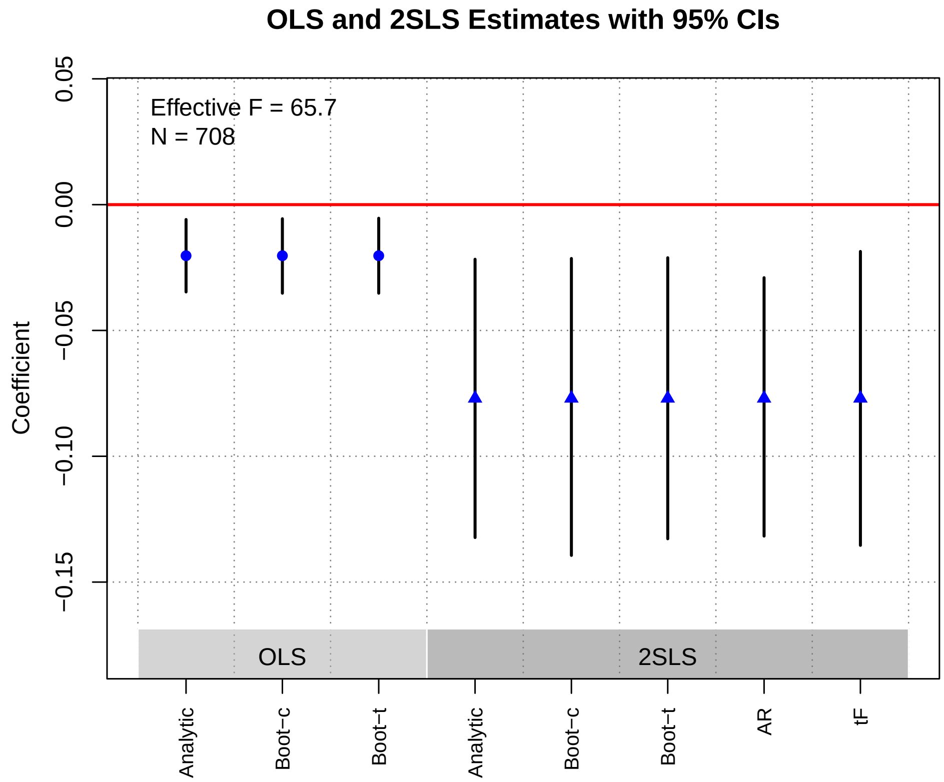
```r
df<-readRDS("./data/ajps_Stokes_2016.rds")
D <-"prop_3km"
Y <- "chng_lib"
Z <- "avg_pwr_log"
controls <- c("mindistlake", "mindistlake_sq", "longitude",
              "long_sq", "latitude", "lat_sq", "long_lat")
cl <- NULL
FE <- "ed_id"
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 31.212 sec.
## AR Test Inversion...

## $est_ols
##             Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -0.0203 0.0073 -2.7638 -0.0347  -0.0059  0.0057
## Boot.c   -0.0203 0.0076 -2.6817 -0.0352  -0.0056  0.0020
## Boot.t   -0.0203 0.0073 -2.7638 -0.0352  -0.0054  0.0060
##
## $est_2sls
##            Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -0.077 0.0282 -2.7289 -0.1323  -0.0217  0.0064
## Boot.c   -0.077 0.0309 -2.4878 -0.1394  -0.0214  0.0080
## Boot.t   -0.077 0.0282 -2.7289 -0.1328  -0.0211  0.0070
##
## $AR
## $AR$Fstat
##        F      df1      df2        p
##   9.8855   1.0000 706.0000   0.0017
##
## $AR$ci.print
## [1] "[-0.1317, -0.0290]"
##
## $AR$ci
## [1] -0.13172512 -0.02902874
##
## $AR$bounded
## [1] TRUE
```

```
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##     67.9032     65.7306          NA     63.5161     65.7306
##
## $rho
## [1] 0.3025
##
## $tF
##       F      cF    Coef      SE       t  CI2.5% CI97.5% p-value
## 65.7306  2.0693 -0.0770  0.0282 -2.7289 -0.1354 -0.0186  0.0097
##
## $est_rf
##                Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## avg_pwr_log -0.0585 0.0216  0.0069 0.0227  -0.1019   -0.0157     0.008
##
## $est_fs
##               Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## avg_pwr_log 0.7602 0.0938       0 0.0954   0.5499    0.9324         0
##
## $p_iv
## [1] 1
##
## $N
## [1] 708
##
## $N_cl
## NULL
##
## $df
## [1] 674
##
## $nvalues
##      chng_lib prop_3km avg_pwr_log
## [1,]      708        2         708
```

```
plot_coef(g)
```



Tajima (2013)

Replication Summary	
Unit of analysis	village and urban neighborhood
Treatment	distance to police posts (as a proxy for exposure to military intervention)
Instrument	distance to health station
Outcome	incidence of communal violence
Model	Table1(4)

```
df<-readRDS("./data/ajps_Tajima_2013.rds")
D <-"z2_distpospol"
Y <- "horiz2"
Z <- "z2_dispuskes"
controls <- c("flat", "z2_altitude","urban", "natres", "z2_logvillpop", "z2_logdensvil",
              "z2_povrateksvil", "z2_fgtksvild", "z2_covyredvil", "z2_npwperhh",
              "z2_ethfractvil","z2_ethfractsd", "z2_ethfractd", "z2_relfractvil",
              "z2_relfractsd", "z2_relfractd", "z2_ethclustsd", "z2_ethclustvd",
              "z2_relclustsd", "z2_relclustvd", "z2_wgcovegvil", "z2_wgcovegsd",
              "z2_wgcovegd", "z2_wgcovrgvil", "z2_wgcovrgsd", "z2_wgcovrgd",
              "natdis","javanese_off_java", "islam", "split_kab03", "split_vil03")
cl <- 'kabid03'
FE <- 'prop'
weights<-"probit_touse_wts03"
```

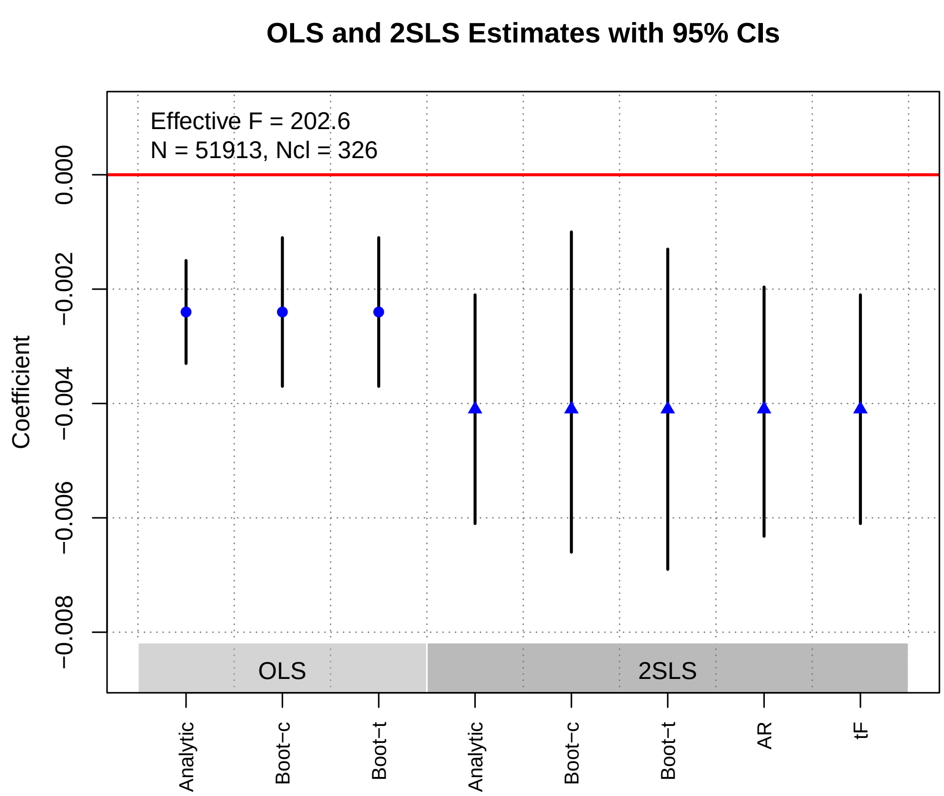
```r
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 1.668 sec.
## AR Test Inversion...
## Parallelising on 15 cores
```

```
## $est_ols
##             Coef    SE       t CI 2.5% CI 97.5% p.value
## Analytic -0.0024 5e-04 -5.2337 -0.0033  -0.0015       0
## Boot.c   -0.0024 7e-04 -3.6186 -0.0037  -0.0011       0
## Boot.t   -0.0024 5e-04 -5.2337 -0.0037  -0.0011       0
##
## $est_2sls
##             Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -0.0041 0.0010 -4.0677 -0.0061  -0.0021   0.000
## Boot.c   -0.0041 0.0014 -2.8886 -0.0066  -0.0010   0.004
## Boot.t   -0.0041 0.0010 -4.0677 -0.0069  -0.0013   0.002
##
## $AR
## $AR$Fstat
##          F        df1        df2          p
##    13.8516     1.0000 51911.0000     0.0002
##
## $AR$ci.print
## [1] "[-0.0063, -0.0020]"
##
## $AR$ci
## [1] -0.006321755 -0.001963142
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##  13363.7649   1529.0807    202.6374    218.7310    202.6374
##
## $rho
## [1] 0.4527
##
## $tF
##        F       cF     Coef       SE         t    CI2.5%   CI97.5%   p-value
## 202.6374   1.9600  -0.0041   0.0010   -4.0677   -0.0061   -0.0021    0.0000
##
```

```
## $est_rf
##                  Coef    SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## z2_dispuskes -0.0019 5e-04       0 6e-04  -0.0029    -5e-04     0.004
##
## $est_fs
##                Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## z2_dispuskes 0.447 0.0114       0 0.0302   0.3895    0.5058         0
##
## $p_iv
## [1] 1
##
## $N
## [1] 51913
##
## $N_cl
## [1] 326
##
## $df
## [1] 51853
##
## $nvalues
##      horiz2 z2_distpospol z2_dispuskes
## [1,]      2           101          101
```

```
plot_coef(g)
```



Trounstine (2016)

Replication Summary	
Unit of analysis	city*year
Treatment	racial segregation
Instrument	the number of waterways in a city; logged population
Outcome	direct general expenditures
Model	Table5(1)

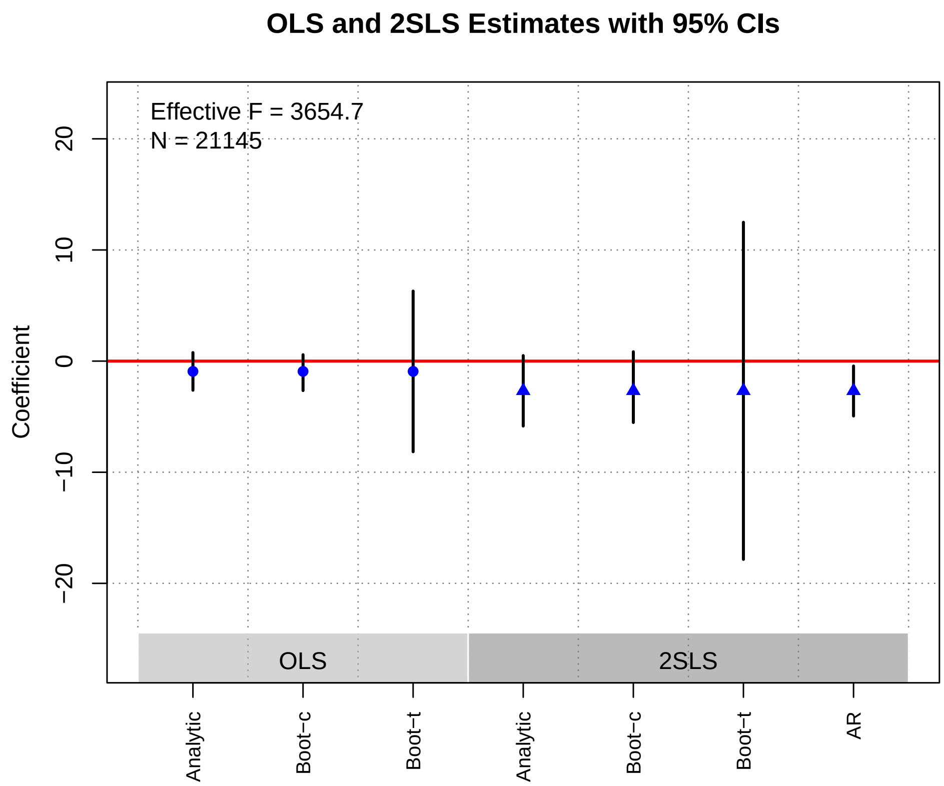
```
df<-readRDS("./data/ajps_Trounstine_2016.rds")
D <-"H_citytract_NHW_i"
Y <- "dgepercap_cpi"
Z <- c("total_rivs_all", "logpop")
controls <- c("dgepercap_cpilag","diversityinterp","pctblkpopinterp",
              "pctasianpopinterp","pctlatinopopinterp","medinciinterp",
              "pctlocalgovworker_100","pctrentersinterp","pctover65",
              "pctcollegegradinterp","northeast","south","midwest",
              "y5", "y6", "y7", "y8", "y9")
cl <- NULL
FE <- NULL
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 25.200 sec.
## AR Test Inversion...
## Parallelising on 15 cores
```

```
## $est_ols
##              Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -0.9265 0.8648 -1.0713 -2.6214   0.7685   0.284
## Boot.c   -0.9265 0.9014 -1.0278 -2.6538   0.5732   0.498
## Boot.t   -0.9265 0.8648 -1.0713 -8.1549   6.3020   0.508
##
## $est_2sls
##              Coef     SE       t  CI 2.5% CI 97.5% p.value
## Analytic -2.6757 1.6174 -1.6543  -5.8458   0.4944  0.0981
## Boot.c   -2.6757 1.7399 -1.5379  -5.5271   0.8405  0.2200
## Boot.t   -2.6757 1.6174 -1.6543 -17.8458  12.4944  0.3030
##
## $AR
## $AR$Fstat
##          F        df1        df2           p
```

```
##     4.1611     2.0000 21142.0000     0.0156
##
## $AR$ci.print
## [1] "[-4.9401, -0.4437]"
##
## $AR$ci
## [1] -4.9400656 -0.4436849
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##    3883.651    2506.495          NA    2574.610    3654.705
##
## $rho
## [1] 0.5185
##
## $est_rf
##                   Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## total_rivs_all -0.0081 0.0229  0.7217 0.0241  -0.0612    0.0267     0.852
## logpop         -0.0855 0.0407  0.0355 0.0437  -0.1565    0.0113     0.132
##
## $est_fs
##                  Coef    SE p.value  SE.b CI.b2.5% CI.b97.5% p.value.b
## total_rivs_all 0.0054 3e-04       0 3e-04   0.0048    0.0060         0
## logpop         0.0291 5e-04       0 5e-04   0.0281    0.0301         0
##
## $p_iv
## [1] 2
##
## $N
## [1] 21145
##
## $N_cl
## NULL
##
## $df
## [1] 21125
##
## $nvalues
##      dgepercap_cpi H_citytract_NHW_i total_rivs_all logpop
## [1,]         21129             15395             22  16223
```

```
plot_coef(g)
```



Vernby (2013)

Replication Summary	
Unit of analysis	municipality*term
Treatment	share of noncitizens in the electorate
Instrument	immigration Inflow 1940–1950; Immigration Inflow 1960–1967
Outcome	municipal education and social spending
Model	Table3(2)

```
df<-readRDS("./data/ajps_Vernby_2013.rds")
D <-"noncitvotsh"
Y <- "Y"
Z <- c("inv1950", "inv1967")
controls <- c("Taxbase2", "L_Taxbase2", "manu", "L_manu", "pop", "L_pop")
cl <- "lan"
FE <- NULL
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

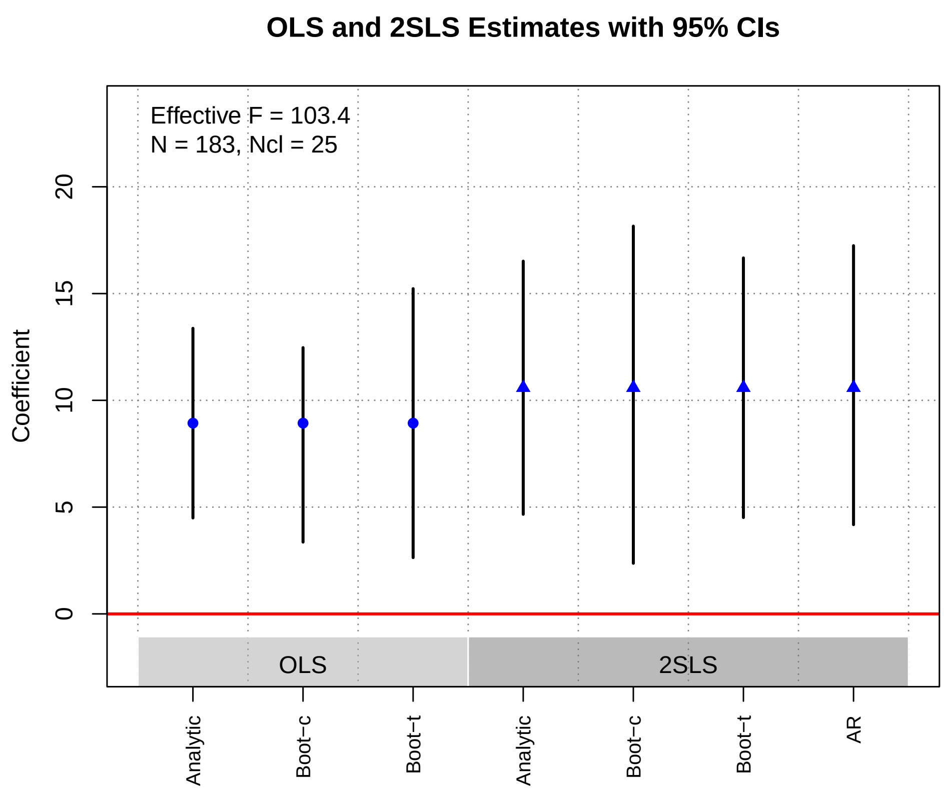
```
## Bootstrapping:
```

```
## Parallelising 1000 reps on 15 cores
## Bootstrap took 12.952 sec.
## AR Test Inversion...

## $est_ols
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 8.9328 2.2655 3.9430  4.4925  13.3731  0.0001
## Boot.c   8.9328 2.3361 3.8238  3.3614  12.4688  0.0000
## Boot.t   8.9328 2.2655 3.9430  2.6345  15.2312  0.0180
##
## $est_2sls
##             Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 10.5903 3.0243 3.5017  4.6626  16.5180   5e-04
## Boot.c   10.5903 4.2880 2.4697  2.3685  18.1580   2e-02
## Boot.t   10.5903 3.0243 3.5017  4.5112  16.6693   1e-03
##
## $AR
## $AR$Fstat
##        F      df1      df2        p
##   7.5357   2.0000 180.0000   0.0007
##
## $AR$ci.print
## [1] "[4.1787, 17.2438]"
##
## $AR$ci
## [1]  4.178671 17.243804
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##     66.2203     49.5670     79.6400     25.8481    103.3586
##
## $rho
## [1] 0.6574
##
## $est_rf
##            Coef     SE p.value    SE.b CI.b2.5% CI.b97.5% p.value.b
## inv1950  2.5029 9.1632  0.7847 12.1395 -24.8339   22.8947     0.910
## inv1967 10.0729 8.4306  0.2322  9.4557  -9.0630   29.2884     0.198
##
## $est_fs
##           Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## inv1950 0.7234 0.3017  0.0165 0.4243  -0.1079    1.5285     0.104
## inv1967 0.4665 0.2878  0.1050 0.3297  -0.2696    0.9822     0.190
```

```
##
## $p_iv
## [1] 2
##
## $N
## [1] 183
##
## $N_cl
## [1] 25
##
## $df
## [1] 175
##
## $nvalues
##         Y noncitvotsh inv1950 inv1967
## [1,] 183         183      25      25
```

```
plot_coef(g)
```



Wood and Grose (2022)

Replication Summary	
Unit of analysis	House member/district
Treatment	incumbent found to have campaign finance violations

Replication Summary	
Instrument	audit
Outcome	legislator Retired
Model	Table2(1)

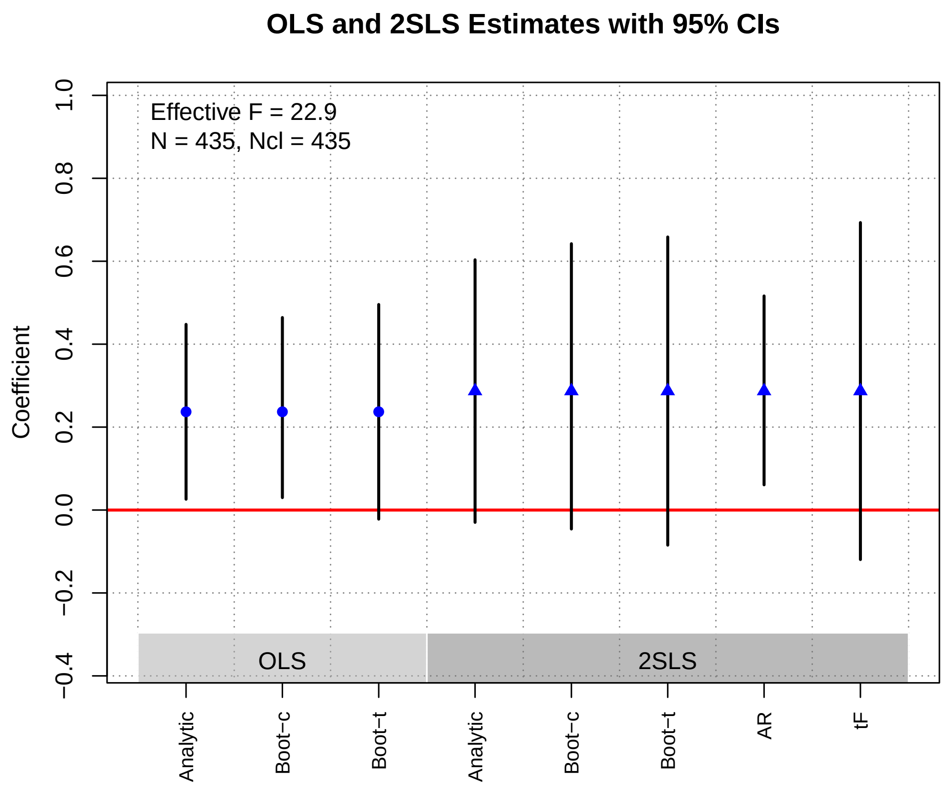
```r
df <-readRDS("./data/ajps_Wood_grose_2022.rds")
## preprocess to generate xwhat and xhat in Stata
D<-"findings"
Y <- "retire__or_resign"
Z <- "audited"
controls <-c("xwhat","south")
cl <- "stcd"
FE <- NULL
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 13.249 sec.
## AR Test Inversion...

## $est_ols
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.2369 0.1076 2.2022  0.0261   0.4477  0.0276
## Boot.c   0.2369 0.1100 2.1527  0.0300   0.4642  0.0200
## Boot.t   0.2369 0.1076 2.2022 -0.0219   0.4956  0.0600
##
## $est_2sls
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.2869 0.1615 1.7764 -0.0297   0.6035  0.0757
## Boot.c   0.2869 0.1710 1.6783 -0.0457   0.6422  0.0860
## Boot.t   0.2869 0.1615 1.7764 -0.0847   0.6585  0.1120
##
## $AR
## $AR$Fstat
##        F      df1      df2        p
##   6.1234   1.0000 433.0000   0.0137
##
## $AR$ci.print
## [1] "[0.0608, 0.5163]"
##
## $AR$ci
## [1] 0.06079225 0.51625575
##
## $AR$bounded
```

```
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##    220.6007     22.8647     22.8647     23.1177     22.8647
##
## $rho
## [1] 0.5819
##
## $tF
##       F      cF    Coef      SE       t  CI2.5% CI97.5% p-value
## 22.8647  2.5155  0.2869  0.1615  1.7764 -0.1194  0.6932  0.1663
##
## $est_rf
##           Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## audited 0.1377 0.0816  0.0916 0.0825  -0.0149    0.3121     0.086
##
## $est_fs
##         Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## audited 0.48 0.1004       0 0.0998     0.28    0.6819         0
##
## $p_iv
## [1] 1
##
## $N
## [1] 435
##
## $N_cl
## [1] 435
##
## $df
## [1] 431
##
## $nvalues
##      retire__or_resign findings audited
## [1,]                 2        2       2
```

```
plot_coef(g)
```



Zhu (2017)

Replication Summary	
Unit of analysis	province*period
Treatment	MNC activity
Instrument	weighted geographic closeness
Outcome	corruption
Model	Table1(1)

```
df <- readRDS("./data/ajps_Zhu_2017.rds")
D <-"MNC"
Y <- "corruption1"
Z <- "lwdist"
controls <- c("lgdpcap6978", "gdp6978", "population", "lgovtexp9302",
              "pubempratio", "leduc", "pwratio", "female", "time")
cl <- NULL
FE <- NULL
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

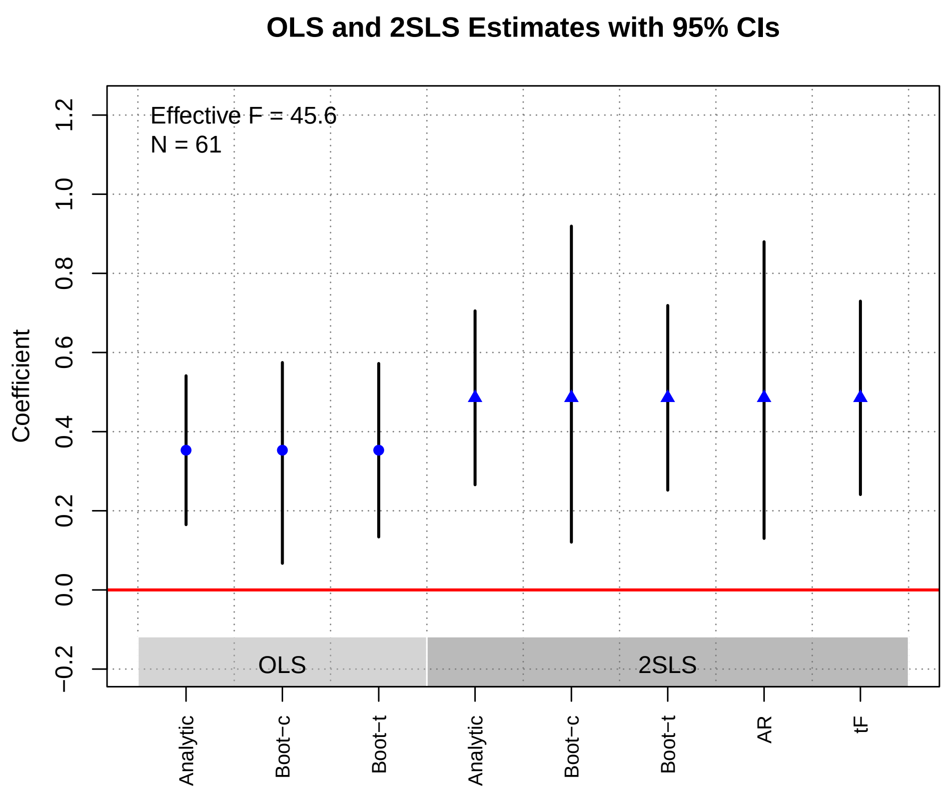
```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 13.058 sec.
```

```
## AR Test Inversion...

## $est_ols
##            Coef    SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.3531 0.096 3.6788  0.1650   0.5412  0.0002
## Boot.c   0.3531 0.125 2.8253  0.0673   0.5746  0.0180
## Boot.t   0.3531 0.096 3.6788  0.1339   0.5722  0.0030
##
## $est_2sls
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.4855 0.1121 4.3317  0.2658   0.7052   0.000
## Boot.c   0.4855 0.2024 2.3986  0.1208   0.9193   0.018
## Boot.t   0.4855 0.1121 4.3317  0.2522   0.7188   0.000
##
## $AR
## $AR$Fstat
##       F     df1     df2       p
##  7.1568  1.0000 59.0000  0.0096
##
## $AR$ci.print
## [1] "[0.1302, 0.8798]"
##
## $AR$ci
## [1] 0.1302327 0.8798029
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##     45.9155     45.5515          NA     23.0582     45.5515
##
## $rho
## [1] 0.6919
##
## $tF
##       F      cF    Coef      SE       t  CI2.5% CI97.5% p-value
## 45.5515  2.1802  0.4855  0.1121  4.3317  0.2411  0.7298  0.0001
##
## $est_rf
##         Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## lwdist 0.559 0.1698   0.001 0.2822   0.1346    1.3403     0.018
##
## $est_fs
##          Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## lwdist 1.1514 0.1706       0 0.2398   0.7729    1.7417         0
```

```
##
## $p_iv
## [1] 1
##
## $N
## [1] 61
##
## $N_cl
## NULL
##
## $df
## [1] 50
##
## $nvalues
##      corruption1 MNC lwdist
## [1,]          61  61     61
```

```
plot_coef(g)
```



JOP

Acharya et al. (2016)

Replication Summary	
Unit of analysis	county
Treatment	slave proportion in 1860
Instrument	measures of the environmental suitability for growing cotton
Outcome	proportion Democrat
Model	Table2(2)

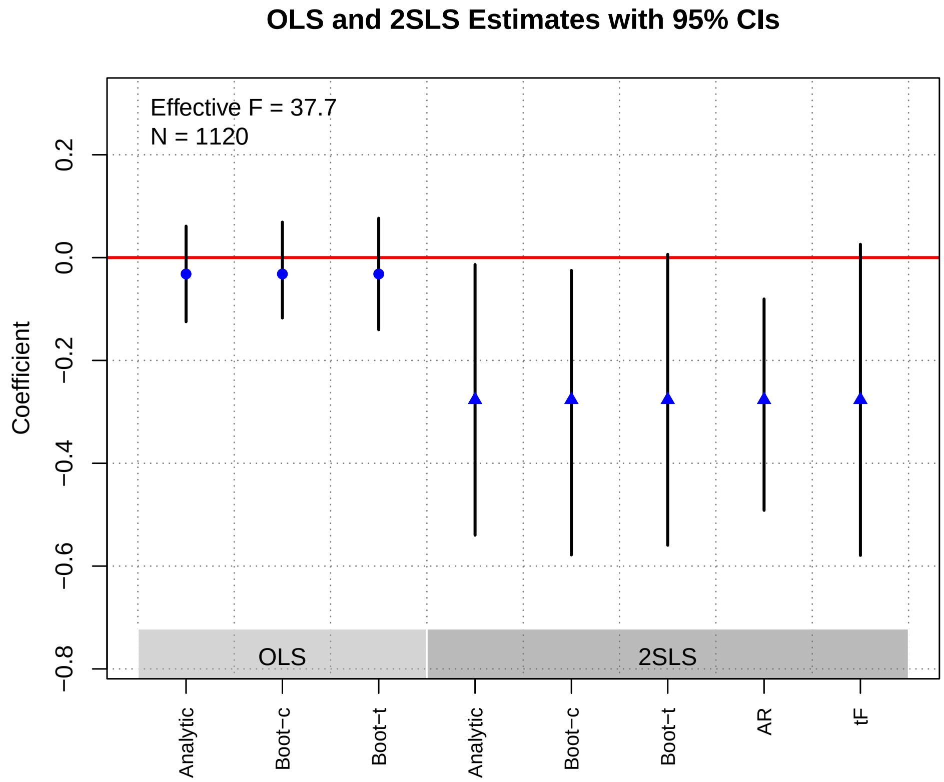
```
df<-readRDS("./data/jop_Acharya_etal_2016.rds")
Y <- "dem"
D <-"pslave1860"
Z <- "cottonsuit"
controls <- c("x2", "rugged", "latitude", "x2", "longitude", "x3","x4", "water1860")
cl <- NULL
FE <- 'code'
weights<-"sample.size"
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 30.430 sec.
## AR Test Inversion...
```

```
## $est_ols
##              Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -0.0318 0.0474 -0.6701 -0.1247   0.0612  0.5028
## Boot.c   -0.0318 0.0473 -0.6716 -0.1174   0.0690  0.5220
## Boot.t   -0.0318 0.0474 -0.6701 -0.1402   0.0766  0.5360
##
## $est_2sls
##              Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -0.2766 0.1343 -2.0596 -0.5399  -0.0134  0.0394
## Boot.c   -0.2766 0.1446 -1.9130 -0.5783  -0.0250  0.0380
## Boot.t   -0.2766 0.1343 -2.0596 -0.5595   0.0063  0.0530
##
## $AR
## $AR$Fstat
##         F       df1       df2         p
##    7.6234    1.0000 1118.0000    0.0059
##
## $AR$ci.print
## [1] "[-0.4915, -0.0805]"
##
## $AR$ci
## [1] -0.4915207 -0.0805310
##
```

```
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##    106.4957     37.6527          NA     34.7203     37.6527
##
## $rho
## [1] 0.2973
##
## $tF
##       F      cF    Coef      SE       t  CI2.5% CI97.5% p-value
## 37.6527  2.2528 -0.2766  0.1343 -2.0596 -0.5792  0.0259  0.0731
##
## $est_rf
##               Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## cottonsuit -0.1128 0.0518  0.0294 0.0548  -0.2156   -0.0103     0.038
##
## $est_fs
##              Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## cottonsuit 0.4079 0.0665       0 0.0692   0.2789    0.5459         0
##
## $p_iv
## [1] 1
##
## $N
## [1] 1120
##
## $N_cl
## NULL
##
## $df
## [1] 1098
##
## $nvalues
##      dem pslave1860 cottonsuit
## [1,] 911       1077       1120
```

```
plot_coef(g)
```



Alt et al. (2016)

Replication Summary	
Unit of analysis	individual
Treatment	unemployment expectations
Instrument	assignment to receiving an aggregate unemployment forecast
Outcome	vote intention
Model	Table2(1)

```
df<- readRDS("./data/jop_Alt_etal_2015.rds")
D <- "urate_fut"
Y <- "gov"
Z <- "treatment"
controls <- "urate_now"
cl <- NULL
FE <- NULL
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

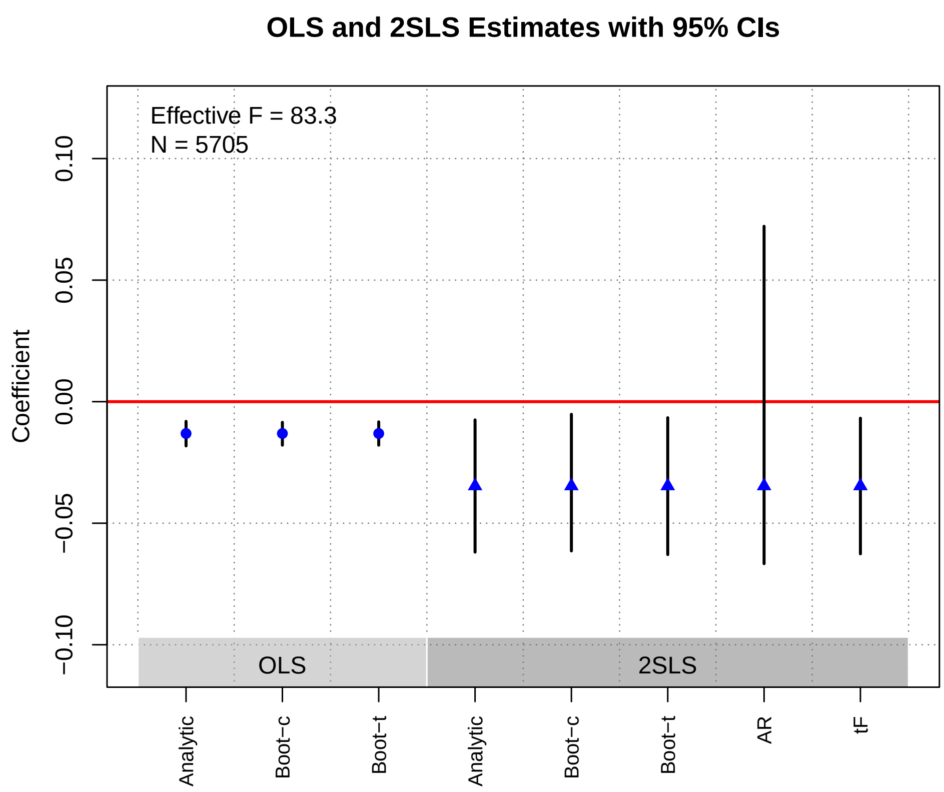
```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 13.488 sec.
```

```
## AR Test Inversion...
## Parallelising on 15 cores

## $est_ols
##             Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -0.0131 0.0026 -5.0845 -0.0182  -0.0081       0
## Boot.c   -0.0131 0.0026 -5.1170 -0.0179  -0.0085       0
## Boot.t   -0.0131 0.0026 -5.0845 -0.0179  -0.0083       0
##
## $est_2sls
##             Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -0.0347 0.0139 -2.5022 -0.0619  -0.0075  0.0123
## Boot.c   -0.0347 0.0144 -2.4149 -0.0614  -0.0052  0.0100
## Boot.t   -0.0347 0.0139 -2.5022 -0.0629  -0.0066  0.0140
##
## $AR
## $AR$Fstat
##         F       df1       df2         p
##    0.0017    1.0000 5703.0000    0.9672
##
## $AR$ci.print
## [1] "[-0.0666, 0.0721]"
##
## $AR$ci
## [1] -0.06664959  0.07214055
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##     60.1863     68.9098          NA     64.7231     83.3152
##
## $rho
## [1] 0.0801
##
## $tF
##       F      cF    Coef      SE       t  CI2.5% CI97.5% p-value
## 83.3152  2.0100 -0.0347  0.0139 -2.5022 -0.0626 -0.0068  0.0147
##
## $est_rf
##            Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## treatment 0.027 0.0243  0.2661 0.0244  -0.0218    0.0759      0.27
##
## $est_fs
##              Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
```

```
## treatment -0.9354 0.1169       0 0.1163  -1.1617   -0.7045         0
##
## $p_iv
## [1] 1
##
## $N
## [1] 5705
##
## $N_cl
## NULL
##
## $df
## [1] 5702
##
## $nvalues
##      gov urate_fut treatment
## [1,]   2        88         8
```

```
plot_coef(g)
```



Arias and Stasavage (2019)

Replication Summary	
Unit of analysis	country*year
Treatment	government expenditures
Instrument	trade shock × UK bond yield
Outcome	regular leader turnover
Model	Table3(2)

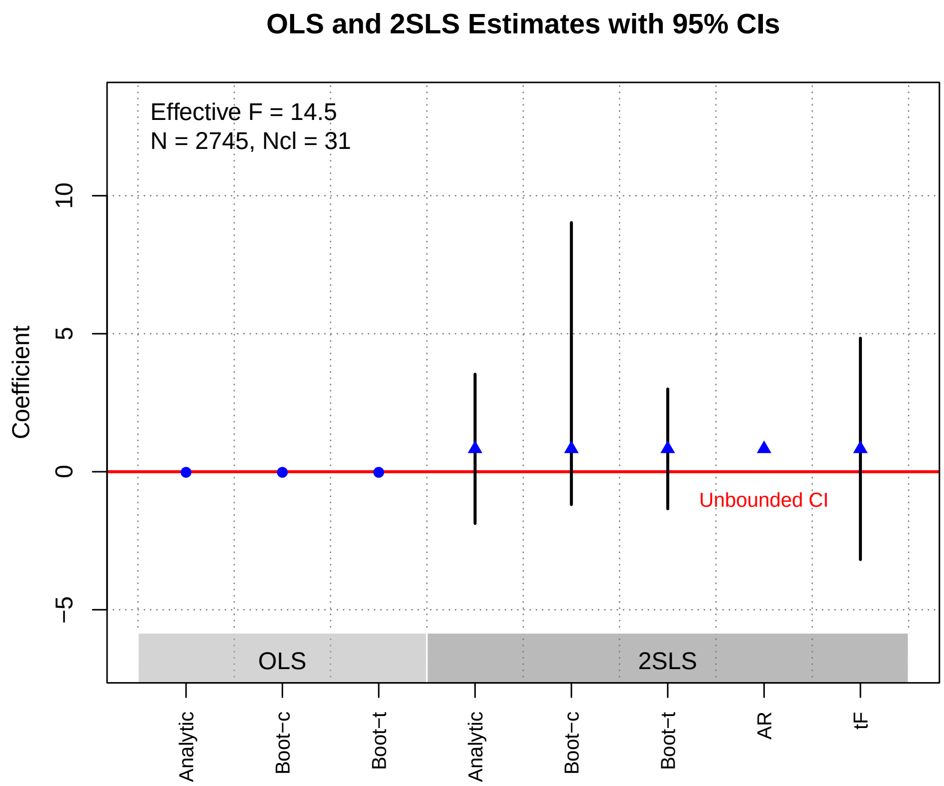
```
# Variables are already residualized against controls, fixed effects, and unit-specific trends
df<-readRDS("./data/jop_Arias_etal_2019.rds")
Y <- "regular_res"
D <- "dexpenditures_res"
Z <- "interact_res"
controls <- NULL
cl<-c("ccode","year")
FE<-NULL
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 14.650 sec.
## AR Test Inversion...
```

```
## $est_ols
##             Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -0.0215 0.0389 -0.5525 -0.0977   0.0547  0.5806
## Boot.c   -0.0215 0.0407 -0.5274 -0.1066   0.0556  0.5551
## Boot.t   -0.0215 0.0389 -0.5525 -0.0977   0.0548  0.5667
##
## $est_2sls
##            Coef      SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.8282  1.3792 0.6005 -1.8749   3.5314  0.5482
## Boot.c   0.8282 80.3169 0.0103 -1.1894   9.0263  0.4936
## Boot.t   0.8282  1.3792 0.6005 -1.3416   2.9981  0.4852
##
## $AR
## $AR$Fstat
##         F       df1       df2         p
##    0.3982    1.0000 2743.0000    0.5281
##
## $AR$ci.print
## [1] "(-Inf, Inf)"
##
## $AR$ci
## [1] -Inf  Inf
```

```
##
## $AR$bounded
## [1] FALSE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##      3.0429      3.4739     14.4763      7.9802     14.4763
##
## $rho
## [1] 0.0333
##
## $tF
##       F      cF    Coef      SE       t  CI2.5% CI97.5% p-value
## 14.4763  2.9071  0.8282  1.3792  0.6005 -3.1812  4.8376  0.6856
##
## $est_rf
##               Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## interact_res 0.276 0.4337  0.5245 0.4674  -0.3928    1.4277    0.4703
##
## $est_fs
##                Coef     SE p.value  SE.b CI.b2.5% CI.b97.5% p.value.b
## interact_res 0.3332 0.1788  0.0623 0.118   0.0787    0.5408    0.0233
##
## $p_iv
## [1] 1
##
## $N
## [1] 2745
##
## $N_cl
## [1] 31
##
## $df
## [1] 2743
##
## $nvalues
##      regular_res dexpenditures_res interact_res
## [1,]        2745              2745         2745
```

```
plot_coef(g)
```



Bhavnani and Lee (2018)

Replication Summary	
Unit of analysis	district*period
Treatment	bureaucrats' embeddedness
Instrument	early-career job assignment
Outcome	proportion of villages with high schools
Model	Table1(4)

```
df <-readRDS("./data/jop_Bhavnani_etal_2018.rds")
D <- "ALLlocal"
Y <- "Phigh"
Z <- "EXALLlocal"
controls <- c("ALLbachdivi", "lnnewpop", "lnnvill", "p_rural", "p_work",
              "p_aglab", "p_sc", "p_st", "lnmurderpc", "stategov", "natgov")
cl <- "distcode71"
FE<- c('distcode71',"year")
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

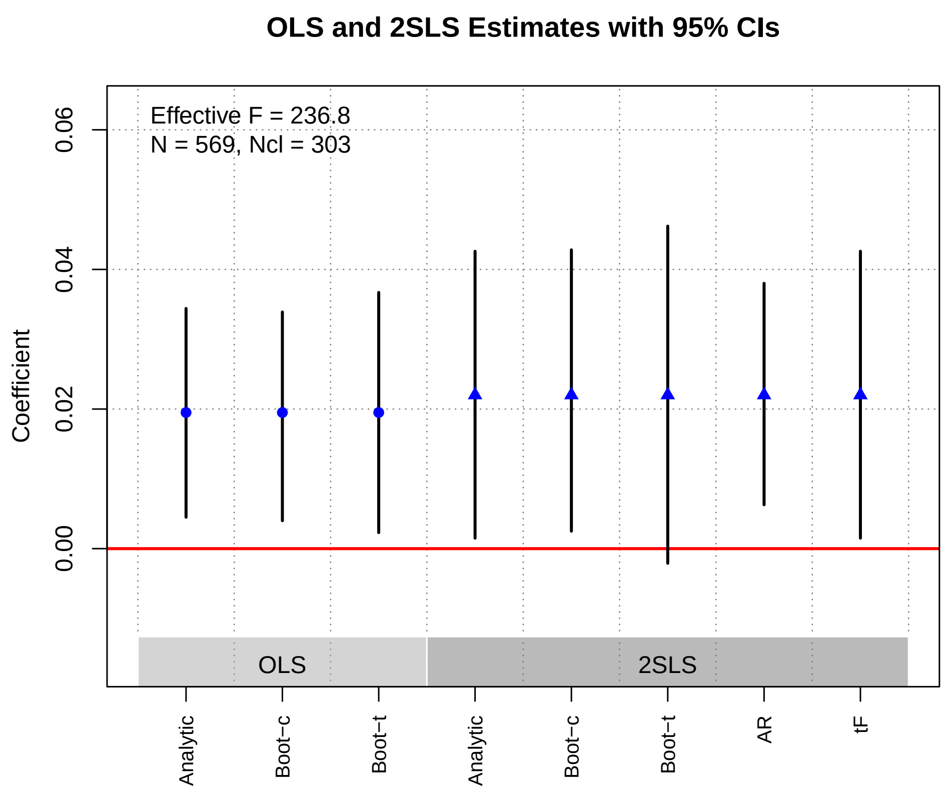
```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 30.052 sec.
```

```
## AR Test Inversion...

## $est_ols
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.0195 0.0076 2.5542  0.0045   0.0344  0.0106
## Boot.c   0.0195 0.0075 2.5996  0.0040   0.0339  0.0200
## Boot.t   0.0195 0.0076 2.5542  0.0023   0.0367  0.0320
##
## $est_2sls
##           Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.022 0.0105 2.0990  0.0015   0.0426  0.0358
## Boot.c   0.022 0.0101 2.1866  0.0025   0.0428  0.0280
## Boot.t   0.022 0.0105 2.0990 -0.0021   0.0462  0.0700
##
## $AR
## $AR$Fstat
##        F      df1      df2        p
##   7.3827   1.0000 567.0000   0.0068
##
## $AR$ci.print
## [1] "[0.0063, 0.0380]"
##
## $AR$ci
## [1] 0.00629017 0.03800371
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##    243.2947    215.8574    236.8206    241.5315    236.8206
##
## $rho
## [1] 0.7002
##
## $tF
##        F       cF     Coef       SE        t   CI2.5%  CI97.5%  p-value
## 236.8206   1.9600   0.0220   0.0105   2.0990   0.0015   0.0426   0.0358
##
## $est_rf
##              Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## EXALLlocal 0.0121 0.0057  0.0344 0.0055   0.0014    0.0234     0.028
##
## $est_fs
##              Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## EXALLlocal 0.5504 0.0375       0 0.0354   0.4828    0.6193         0
```

```
##
## $p_iv
## [1] 1
##
## $N
## [1] 569
##
## $N_cl
## [1] 303
##
## $df
## [1] 253
##
## $nvalues
##      Phigh ALLlocal EXALLlocal
## [1,]   567      493        318
```

```
plot_coef(g)
```



Charron and Lapuente (2013)

Replication Summary	
Unit of analysis	region
Treatment	clientelism

Replication Summary	
Instrument	consolidation of clientelistic networks in regions where rulers have historically less constraints to their decisions
Outcome	quality of governments
Model	Table3(2a)

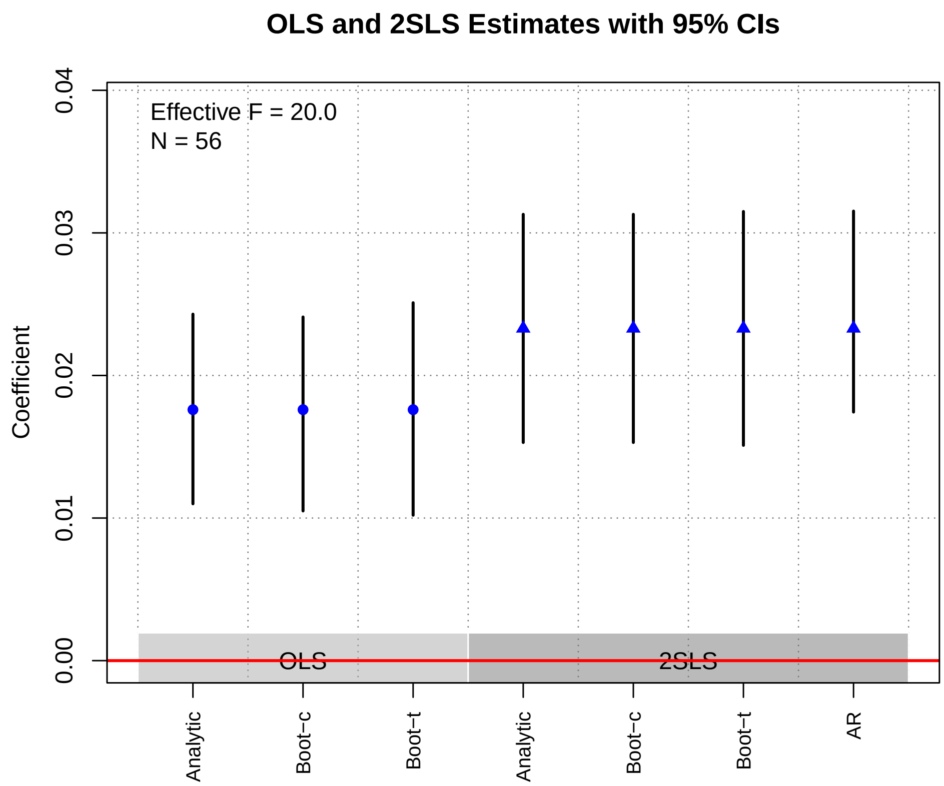
```
df<-readRDS("./data/jop_Charron_etal_2013.rds")
D <- "pc_all4_tol"
Y <- "eqi"
Z <- c("pc_institutions","literacy1880")
controls <- c("logpop", "capitalregion", "ger", "it", "uk","urb_1860_1850_30")
cl <- NULL
FE <- NULL
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 13.168 sec.
## AR Test Inversion...
```

```
## $est_ols
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.0176 0.0034 5.1860  0.0110   0.0243   0.000
## Boot.c   0.0176 0.0035 5.1034  0.0105   0.0241   0.000
## Boot.t   0.0176 0.0034 5.1860  0.0102   0.0251   0.001
##
## $est_2sls
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.0233 0.0041 5.7196  0.0153   0.0313       0
## Boot.c   0.0233 0.0043 5.4129  0.0153   0.0313       0
## Boot.t   0.0233 0.0041 5.7196  0.0151   0.0315       0
##
## $AR
## $AR$Fstat
##       F     df1     df2       p
## 24.8273  2.0000 53.0000  0.0000
##
## $AR$ci.print
## [1] "[0.0174, 0.0315]"
##
## $AR$ci
## [1] 0.01743370 0.03152853
##
## $AR$bounded
```

```
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##     37.2005     31.2712          NA     31.0705     19.9514
##
## $rho
## [1] 0.7828
##
## $est_rf
##                   Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## pc_institutions 0.1941 0.0765  0.0111 0.0802   0.0381    0.3489      0.01
## literacy1880    0.0204 0.0043  0.0000 0.0047   0.0102    0.0294      0.00
##
## $est_fs
##                    Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## pc_institutions 12.1093 2.3469   0e+00 2.4777   7.5947   17.1954     0.000
## literacy1880     0.5348 0.1319   1e-04 0.1525   0.1837    0.7807     0.008
##
## $p_iv
## [1] 2
##
## $N
## [1] 56
##
## $N_cl
## NULL
##
## $df
## [1] 48
##
## $nvalues
##      eqi pc_all4_tol pc_institutions literacy1880
## [1,]  56          44              14           38
```

```
plot_coef(g)
```



Charron et al. (2017)

Replication Summary	
Unit of analysis	region
Treatment	more developed bureaucracy
Instrument	proportion of Protestant residents in a region; aggregate literacy in 1880
Outcome	percent of single bidders in procurement contracts
Model	Table5(4)

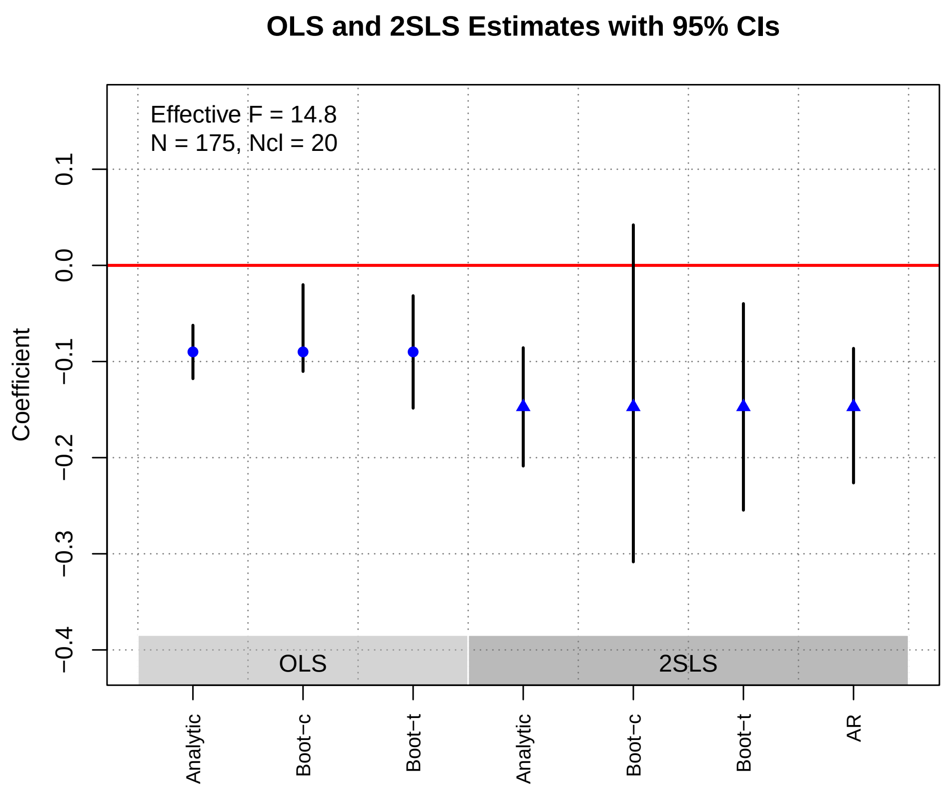
```
df <- readRDS("./data/jop_Charron_etal_2017.rds")
D <- "pubmerit"
Y <- "lcri_euc1_r"
Z <- c("litrate_1880", 'pctprot')
controls <- c("logpopdens", "logppp11", "trust", "pctwomenparl")
cl <- "country"
FE <- NULL
weights<-"eu_popweights"
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 13.236 sec.
## AR Test Inversion...
```

```
## $est_ols
##           Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -0.09 0.0142 -6.3571 -0.1178  -0.0623   0.000
## Boot.c   -0.09 0.0230 -3.9171 -0.1103  -0.0201   0.006
## Boot.t   -0.09 0.0142 -6.3571 -0.1485  -0.0316   0.008
##
## $est_2sls
##             Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -0.1472 0.0314 -4.6940 -0.2087  -0.0857    0.00
## Boot.c   -0.1472 0.0904 -1.6293 -0.3084   0.0421    0.10
## Boot.t   -0.1472 0.0314 -4.6940 -0.2546  -0.0398    0.02
##
## $AR
## $AR$Fstat
##        F      df1      df2        p
##  15.4142   2.0000 172.0000   0.0000
##
## $AR$ci.print
## [1] "[-0.2263, -0.0864]"
##
## $AR$ci
## [1] -0.22625840 -0.08637539
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##     27.8775     23.2292     36.2651      6.7931     14.8219
##
## $rho
## [1] 0.4992
##
## $est_rf
##                 Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## litrate_1880 -0.0009 0.0003  0.0036 0.0006  -0.0019    0.0005     0.222
## pctprot      -0.1769 0.0687  0.0100 0.1433  -0.4448    0.1249     0.312
##
## $est_fs
##                Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## litrate_1880 0.0060 0.0016   2e-04 0.0030  -0.0003    0.0118     0.064
## pctprot      1.1959 0.2723   0e+00 0.4654  -0.0038    1.9179     0.052
##
## $p_iv
## [1] 2
##
```

```
## $N
## [1] 175
##
## $N_cl
## [1] 20
##
## $df
## [1] 169
##
## $nvalues
##      lcri_euc1_r pubmerit litrate_1880 pctprot
## [1,]         173      173           78     131
```

```
plot_coef(g)
```



Cirone and Van Coppenolle (2018)

Replication Summary	
Unit of analysis	deputy*year
Treatment	budget committee service
Instrument	random assignment of budget incumbents to bureaux
Outcome	legislator sponsorship on a budget bill
Model	Table2(2)

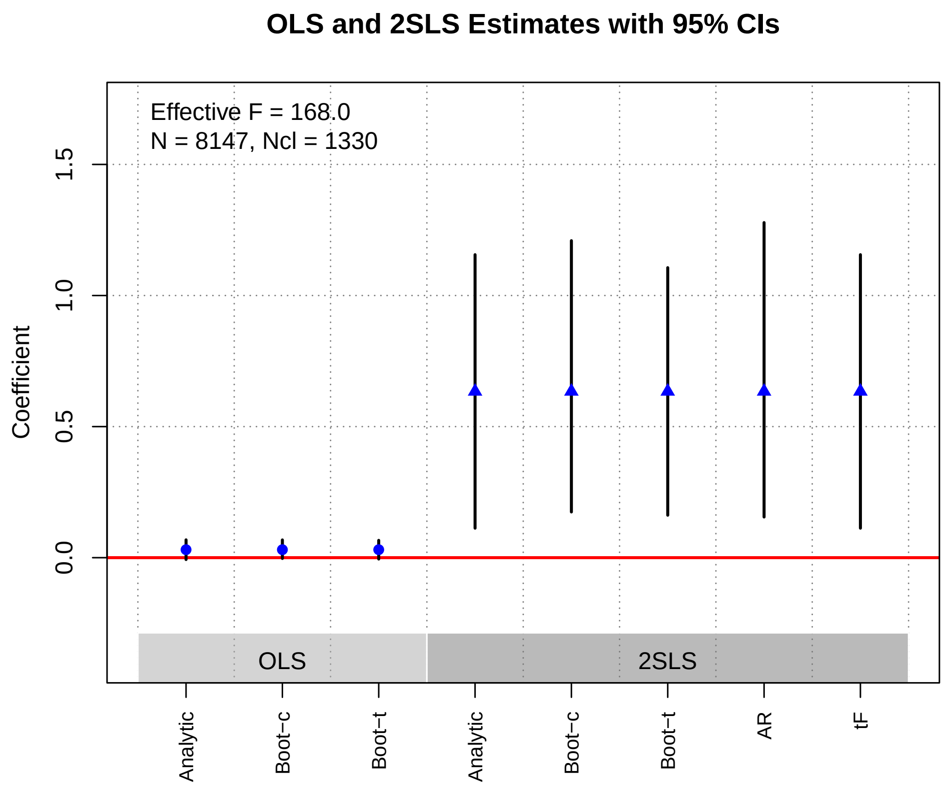
```r
df<- readRDS("./data/jop_Cirone_etal_2018.rds")
D <- "budget"
Y <- "F1to5billbudgetdummy"
Z <- "bureauotherbudgetincumbent"
controls <- c("budgetincumbent", "cummyears", "cummyears2",
              "age", "age2", "permargin", "permargin2",
              "inscrits", "inscrits2", "proprietaire",
              "lib_all", "civil", "paris")
cl <- c("id","year")
FE <- "year"
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 41.643 sec.
## AR Test Inversion...
## Parallelising on 15 cores
```

```
## $est_ols
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.0305 0.0192 1.5883 -0.0071   0.0681  0.1122
## Boot.c   0.0305 0.0179 1.7018 -0.0030   0.0678  0.0720
## Boot.t   0.0305 0.0192 1.5883 -0.0052   0.0661  0.0890
##
## $est_2sls
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.6341 0.2661 2.3827  0.1125   1.1557  0.0172
## Boot.c   0.6341 0.2653 2.3900  0.1744   1.2092  0.0060
## Boot.t   0.6341 0.2661 2.3827  0.1620   1.1062  0.0120
##
## $AR
## $AR$Fstat
##         F       df1       df2         p
##    6.6805    1.0000 8145.0000    0.0098
##
## $AR$ci.print
## [1] "[0.1551, 1.2781]"
##
## $AR$ci
## [1] 0.1550666 1.2781466
##
## $AR$bounded
## [1] TRUE
##
##
```

```
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##     32.1302     34.2557    168.0023     31.4578    168.0023
##
## $rho
## [1] 0.0628
##
## $tF
##        F       cF     Coef       SE        t   CI2.5%  CI97.5%  p-value
## 168.0023   1.9600   0.6341   0.2661   2.3827   0.1125   1.1557   0.0172
##
## $est_rf
##                               Coef    SE p.value  SE.b CI.b2.5% CI.b97.5%
## bureauotherbudgetincumbent -0.0052 0.002  0.0099 0.002  -0.0092   -0.0014
##                            p.value.b
## bureauotherbudgetincumbent     0.006
##
## $est_fs
##                               Coef     SE p.value   SE.b CI.b2.5% CI.b97.5%
## bureauotherbudgetincumbent -0.0083 0.0014       0 0.0015  -0.0113   -0.0055
##                            p.value.b
## bureauotherbudgetincumbent         0
##
## $p_iv
## [1] 1
##
## $N
## [1] 8147
##
## $N_cl
## [1] 1330
##
## $df
## [1] 13
##
## $nvalues
##      F1to5billbudgetdummy budget bureauotherbudgetincumbent
## [1,]                    2      2                          9
```

```
plot_coef(g)
```



Dietrich and Wright (2015)

Replication Summary	
Unit of analysis	transition
Treatment	economic aid
Instrument	constructed Z
Outcome	transitions to multipartyism
Model	Table1(2)

```
df <- readRDS("./data/jop_Dietrich_2015.rds")
D <- "econaid"
Y <- "mp"
Z <- c("Iinfl3","econaid_lgdp_g", "econaid_lpop_g",
       "econaid_cwar_g", "econaid_dnmp_g",
       "econaid_dnmp2_g", "econaid_dnmp3_g")
controls <- c('lgdp', 'lpop', 'cwar', 'dmp',
              'dmp2', 'dmp3', "dnmp", "dnmp2", "dnmp3")
cl<- "cowcode"
FE <- NULL
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

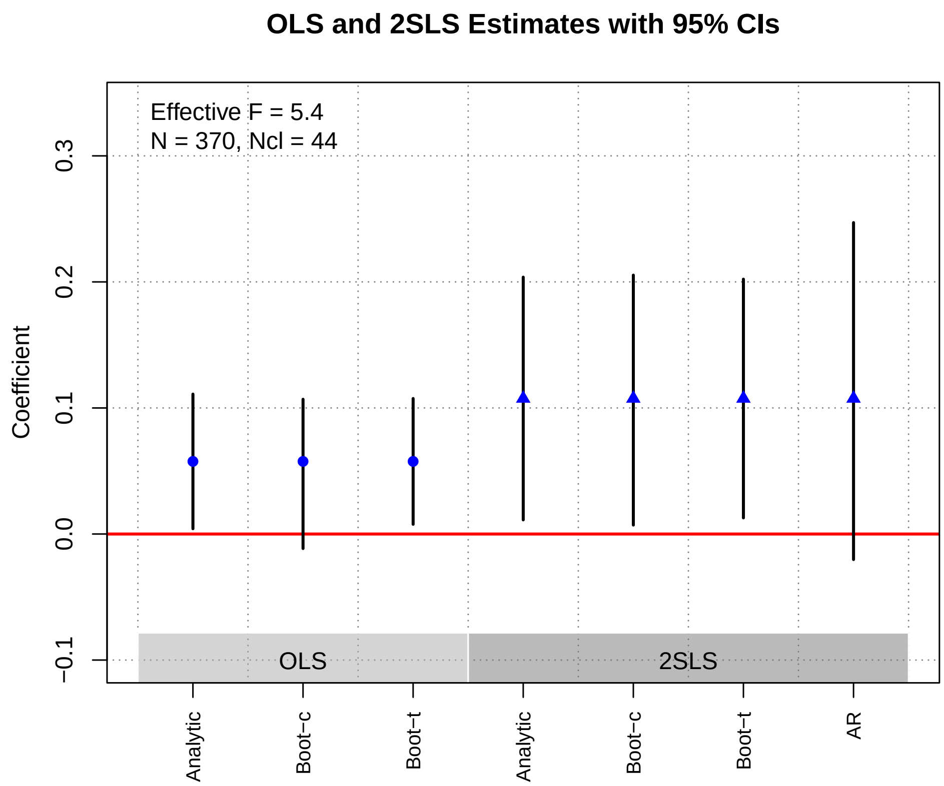
```
## Bootstrapping:
```

```
## Parallelising 1000 reps on 15 cores
## Bootstrap took 13.678 sec.
## AR Test Inversion...

## $est_ols
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.0576 0.0272 2.1133  0.0042   0.1110  0.0346
## Boot.c   0.0576 0.0291 1.9802 -0.0115   0.1069  0.0800
## Boot.t   0.0576 0.0272 2.1133  0.0077   0.1075  0.0230
##
## $est_2sls
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.1075 0.0491 2.1878  0.0112   0.2038  0.0287
## Boot.c   0.1075 0.0485 2.2182  0.0071   0.2054  0.0420
## Boot.t   0.1075 0.0491 2.1878  0.0128   0.2022  0.0280
##
## $AR
## $AR$Fstat
##        F      df1      df2        p
##   1.6483   7.0000 362.0000   0.1207
##
## $AR$ci.print
## [1] "[-0.0203, 0.2471]"
##
## $AR$ci
## [1] -0.02025489  0.24706528
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##     28.9900     47.6878     22.5931      2.1661      5.4068
##
## $rho
## [1] 0.6026
##
## $est_rf
##                    Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## Iinfl3           0.0382 0.0166  0.0214 0.0225  -0.0150    0.0731     0.176
## econaid_lgdp_g   0.0459 0.0330  0.1647 0.0527   0.0061    0.2177     0.032
## econaid_lpop_g   0.0049 0.0253  0.8469 0.0369  -0.0468    0.1032     0.756
## econaid_cwar_g  -0.0084 0.0733  0.9086 0.1026  -0.2346    0.1595     0.872
## econaid_dnmp_g  -0.0227 0.0262  0.3853 0.0323  -0.0801    0.0470     0.584
## econaid_dnmp2_g  0.0010 0.0012  0.3965 0.0015  -0.0021    0.0034     0.638
## econaid_dnmp3_g  0.0000 0.0000  0.4922 0.0000   0.0000    0.0000     0.746
```

```
##
## $est_fs
##                     Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## Iinfl3            0.1561 0.0298  0.0000 0.0607   0.0003    0.2310     0.050
## econaid_lgdp_g    0.1664 0.0605  0.0059 0.2606  -0.4516    0.6285     0.550
## econaid_lpop_g    0.1839 0.0450  0.0000 0.1627  -0.2910    0.3743     0.328
## econaid_cwar_g   -0.2848 0.1453  0.0501 0.5659  -1.8627    0.3582     0.456
## econaid_dnmp_g   -0.0235 0.0500  0.6379 0.1002  -0.2548    0.1441     0.792
## econaid_dnmp2_g  -0.0009 0.0024  0.7103 0.0051  -0.0090    0.0111     0.922
## econaid_dnmp3_g   0.0000 0.0000  0.2738 0.0001  -0.0001    0.0001     0.738
##
## $p_iv
## [1] 7
##
## $N
## [1] 370
##
## $N_cl
## [1] 44
##
## $df
## [1] 362
##
## $nvalues
##      mp econaid Iinfl3 econaid_lgdp_g econaid_lpop_g econaid_cwar_g
## [1,]  2     370    370            370            370            370
##      econaid_dnmp_g econaid_dnmp2_g econaid_dnmp3_g
## [1,]            370             370             370
```

```
plot_coef(g)
```



DiGiuseppe and Shea (2022)

Replication Summary	
Unit of analysis	country*year
Treatment	US support
Instrument	echelon corridor
Outcome	property rights
Model	Table1(5)

```
df <-readRDS("./data/jop_digiuseppe_2022.rds")
D <- "wi_usa_median"
Y<-"Fwi_v2stfisccap2"
Z <- "Echelon2"
controls <-c("wi_v2xcl_prpty","wi_compete", "wi_lnpop_wdi",
             "wi_lngdppc", "wi_polity2", "wi_polity2_2",  "wi_ny_gdp_totl_rt_zs",
             "wi_cwyrs", "wi_c2", "wi_c3", "coldwar")
cl<- NULL
FE<- NULL
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

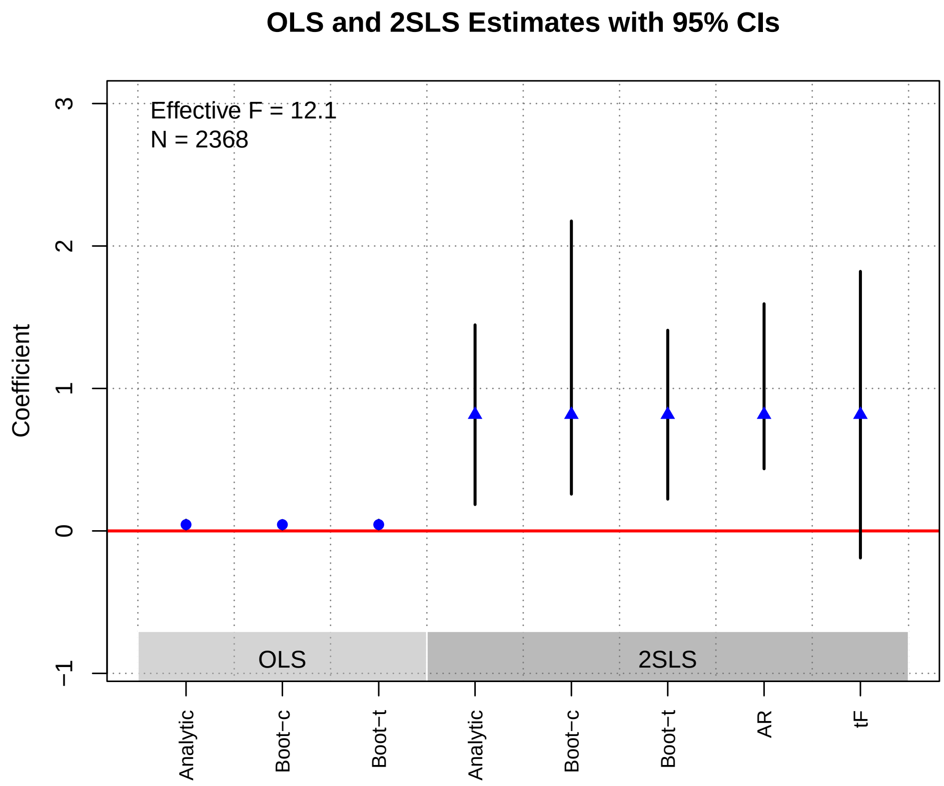
```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
```

```
## Bootstrap took 13.713 sec.
## AR Test Inversion...

## $est_ols
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.0443 0.0156 2.8331  0.0136   0.0749  0.0046
## Boot.c   0.0443 0.0157 2.8182  0.0134   0.0731  0.0080
## Boot.t   0.0443 0.0156 2.8331  0.0137   0.0748  0.0040
##
## $est_2sls
##            Coef     SE     t CI 2.5% CI 97.5% p.value
## Analytic 0.8158 0.3217 2.536  0.1853   1.4463  0.0112
## Boot.c   0.8158 0.6373 1.280  0.2583   2.1753  0.0040
## Boot.t   0.8158 0.3217 2.536  0.2232   1.4085  0.0110
##
## $AR
## $AR$Fstat
##         F       df1       df2         p
##   21.8229    1.0000 2366.0000    0.0000
##
## $AR$ci.print
## [1] "[0.4362, 1.5943]"
##
## $AR$ci
## [1] 0.4362239 1.5943165
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##     18.8218     12.1084          NA     12.2382     12.1084
##
## $rho
## [1] 0.089
##
## $tF
##       F      cF    Coef      SE       t  CI2.5% CI97.5% p-value
## 12.1084  3.1262  0.8158  0.3217  2.5360 -0.1899  1.8215  0.1118
##
## $est_rf
##            Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## Echelon2 0.1792 0.0615  0.0036 0.0623   0.0582    0.3015     0.004
##
## $est_fs
##            Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
```

```
## Echelon2 0.2196 0.0631   5e-04 0.0628   0.0865    0.3336         0
##
## $p_iv
## [1] 1
##
## $N
## [1] 2368
##
## $N_cl
## NULL
##
## $df
## [1] 2355
##
## $nvalues
##      Fwi_v2stfisccap2 wi_usa_median Echelon2
## [1,]              314          2368        2
```

```
plot_coef(g)
```



Dube and Naidu (2015)

Replication Summary	
Unit of analysis	municipality*year
Treatment	changes in US funding to Colombia
Instrument	US funding in countries outside of Latin America
Outcome	the number of paramilitary attacks
Model	Table1(1)

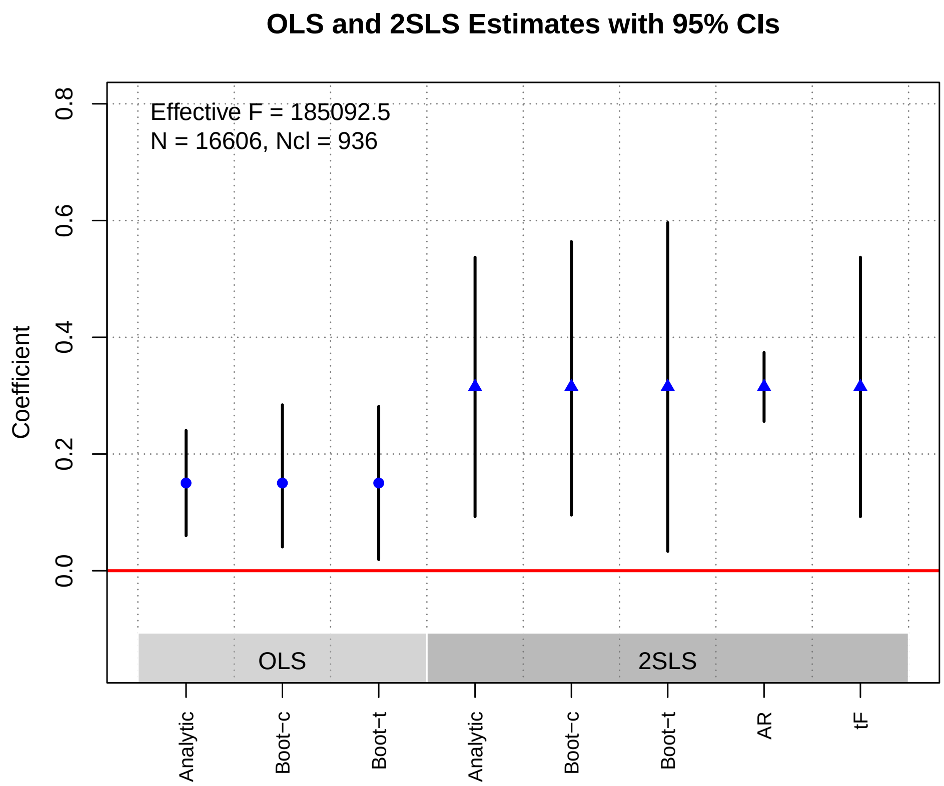
```
df<-readRDS("./data/jop_Dube_etal_2015.rds")
D <- "bases6xlrmilnar_col"
Y <- "paratt"
Z <- "bases6xlrmilwnl"
controls <-"lnnewpop"
cl <- "municipality"
FE <- c("year","municipality")
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 46.482 sec.
## AR Test Inversion...
## Parallelising on 15 cores
```

```
## $est_ols
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.1503 0.0460 3.2692  0.0602   0.2404  0.0011
## Boot.c   0.1503 0.0628 2.3928  0.0408   0.2843  0.0100
## Boot.t   0.1503 0.0460 3.2692  0.0192   0.2815  0.0330
##
## $est_2sls
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.3149 0.1134 2.7771  0.0927   0.5372  0.0055
## Boot.c   0.3149 0.1203 2.6169  0.0954   0.5639  0.0020
## Boot.t   0.3149 0.1134 2.7771  0.0333   0.5965  0.0330
##
## $AR
## $AR$Fstat
##          F        df1        df2          p
##   108.9533     1.0000 16604.0000     0.0000
##
## $AR$ci.print
## [1] "[0.2560, 0.3739]"
##
## $AR$ci
## [1] 0.2559553 0.3738911
```

```
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##   7003.8727    810.8395 185092.5288 175966.6248 185092.5288
##
## $rho
## [1] 0.556
##
## $tF
##           F          cF        Coef          SE           t      CI2.5%
## 185092.5288      1.9600      0.3149      0.1134      2.7771      0.0927
##     CI97.5%     p-value
##      0.5372      0.0055
##
## $est_rf
##                   Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## bases6xlrmilwnl 1.1155 0.3994  0.0052 0.4263   0.3379    1.9989     0.002
##
## $est_fs
##                   Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## bases6xlrmilwnl 3.5422 0.1244       0 0.0084   3.5237    3.5565         0
##
## $p_iv
## [1] 1
##
## $N
## [1] 16606
##
## $N_cl
## [1] 936
##
## $df
## [1] 935
##
## $nvalues
##      paratt bases6xlrmilnar_col bases6xlrmilwnl
## [1,]     13                  19              18
```

```
plot_coef(g)
```



Feigenbaum and Hall (2015)

Replication Summary	
Unit of analysis	congressional district*decade
Treatment	localized trade shocks in congressional districts
Instrument	Chinese exports to other economies*local exposure
Outcome	trade score based on congressional voting
Model	Table1(3)

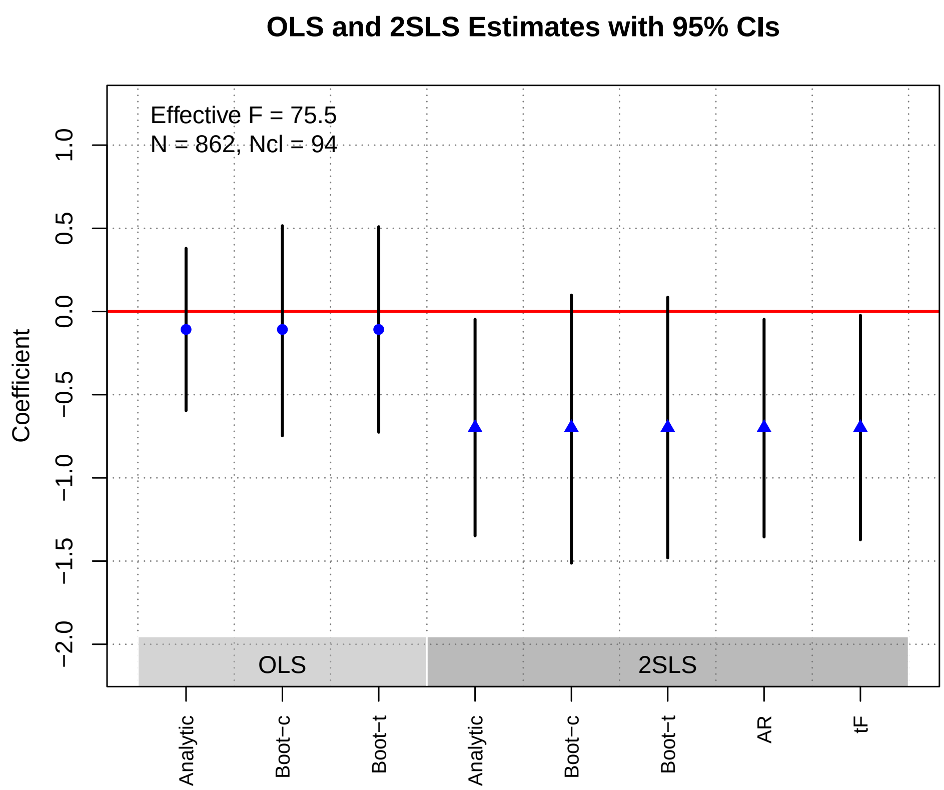
```
df<-readRDS("./data/jop_Feigenbaum_etal_2015.rds")
D <-"x"
Y <- "tradescore"
Z <- "z"
controls <- c("dem_share")
cl <- "state_cluster"
FE <- "decade"
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 28.519 sec.
## AR Test Inversion...
```

```
## $est_ols
##             Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -0.108 0.2489 -0.4339 -0.5958   0.3798  0.6643
## Boot.c   -0.108 0.3238 -0.3335 -0.7464   0.5155  0.7240
## Boot.t   -0.108 0.2489 -0.4339 -0.7255   0.5095  0.7150
##
## $est_2sls
##              Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -0.6976 0.3321 -2.1004 -1.3486  -0.0466  0.0357
## Boot.c   -0.6976 0.4066 -1.7156 -1.5123   0.0987  0.0840
## Boot.t   -0.6976 0.3321 -2.1004 -1.4809   0.0857  0.0700
##
## $AR
## $AR$Fstat
##        F      df1      df2        p
##   4.4060   1.0000 860.0000   0.0361
##
## $AR$ci.print
## [1] "[-1.3552, -0.0466]"
##
## $AR$ci
## [1] -1.35524026 -0.04662122
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##   1189.3393    204.4798     75.5233     74.1384     75.5233
##
## $rho
## [1] 0.7622
##
## $tF
##       F      cF    Coef      SE       t  CI2.5% CI97.5% p-value
## 75.5233  2.0310 -0.6976  0.3321 -2.1004 -1.3722 -0.0231  0.0427
##
## $est_rf
##      Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## z -0.5863 0.2683  0.0289 0.3461  -1.2788    0.0863     0.084
##
## $est_fs
##     Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## z 0.8405 0.0588       0 0.0976   0.6957    1.0709         0
##
## $p_iv
```

```
## [1] 1
##
## $N
## [1] 862
##
## $N_cl
## [1] 94
##
## $df
## [1] 858
##
## $nvalues
##      tradescore   x   z
## [1,]        709 698 697
```

```
plot_coef(g)
```



Flores-Macias and Kreps (2013)

Replication Summary	
Unit of analysis	country*year
Treatment	trade volume
Instrument	lagged energy production
Outcome	foreign policy convergence

Replication Summary	
Model	Table2(1)

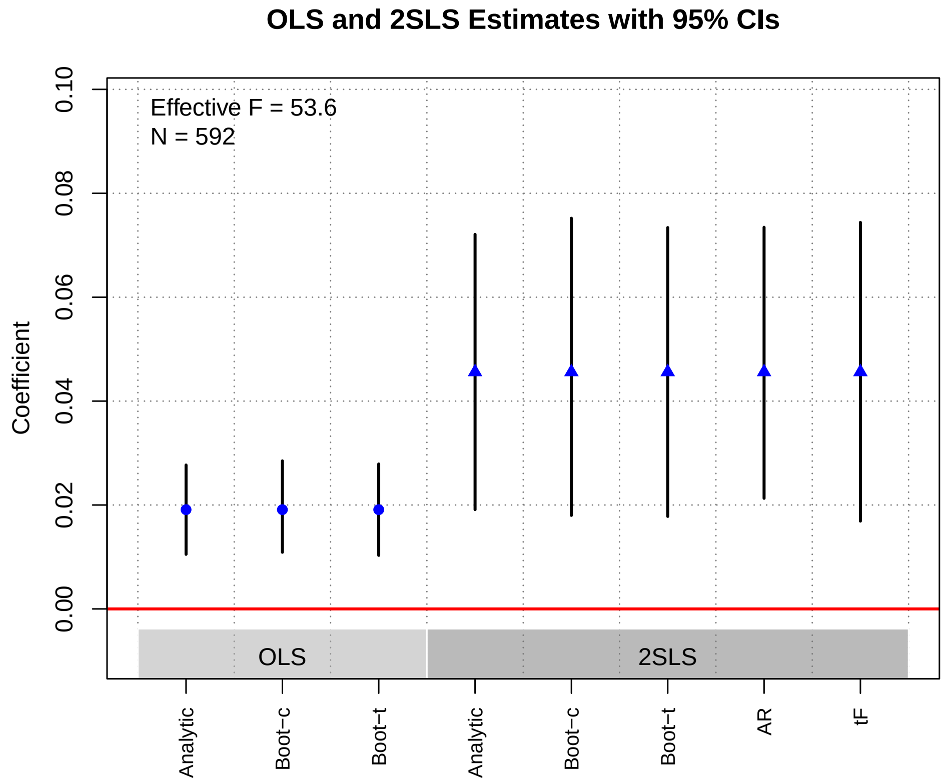
```r
df<- readRDS("./data/jop_Flores_etal_2013.rds")
D <- "log_tot_trade"
Y <- "log_HRVOTE"
Z <- "lag_log_energ_prod"
controls <- c("log_cinc", "us_aid100", "log_tot_ustrade",
              "Joint_Dem_Dum", "pts_score", "dummy2004")
cl <- NULL
FE <- 'statea'
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 28.491 sec.
## AR Test Inversion...

## $est_ols
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.0191 0.0044 4.3531  0.0105   0.0277       0
## Boot.c   0.0191 0.0044 4.3359  0.0109   0.0285       0
## Boot.t   0.0191 0.0044 4.3531  0.0103   0.0279       0
##
## $est_2sls
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.0456 0.0135 3.3747  0.0191   0.0721   7e-04
## Boot.c   0.0456 0.0147 3.1102  0.0180   0.0752   0e+00
## Boot.t   0.0456 0.0135 3.3747  0.0178   0.0734   1e-03
##
## $AR
## $AR$Fstat
##        F      df1      df2        p
##  13.0878   1.0000 590.0000   0.0003
##
## $AR$ci.print
## [1] "[0.0213, 0.0735]"
##
## $AR$ci
## [1] 0.0212854 0.0734630
##
## $AR$bounded
## [1] TRUE
##
```

```
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##     66.1143     53.6345          NA     50.6690     53.6345
##
## $rho
## [1] 0.3295
##
## $tF
##       F      cF    Coef      SE       t  CI2.5% CI97.5% p-value
## 53.6345  2.1276  0.0456  0.0135  3.3747  0.0169  0.0744  0.0019
##
## $est_rf
##                     Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## lag_log_energ_prod 0.1086 0.0301   3e-04 0.0318   0.0438    0.1708         0
##
## $est_fs
##                     Coef    SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## lag_log_energ_prod 2.3803 0.325       0 0.3344   1.7606     3.114         0
##
## $p_iv
## [1] 1
##
## $N
## [1] 592
##
## $N_cl
## NULL
##
## $df
## [1] 543
##
## $nvalues
##      log_HRVOTE log_tot_trade lag_log_energ_prod
## [1,]         32           590                581
```

```
plot_coef(g)
```



Gehlbach and Keefer (2012)

Replication Summary	
Unit of analysis	nondemocratic episode
Treatment	age of ruling party less leader years in office
Instrument	whether the first ruler in a nondemocratic episode is a military leader
Outcome	private invest
Model	Table1(4)

```
df<- readRDS("./data/jop_Gelbach_etal_2012.rds")
D <- "gov1_yrs"
Y <- "gfcf_priv_gdp"
Z <- "military_first_alt"
controls <- c("tenure", "stabs", "fuelex_gdp", "oresex_gdp",
              "frac_ethn", "frac_relig", "frac_ling", "pop_yng_pct",
              "pop_tot", "pop_ru_pct", "land_km", "gdppc_ppp_2005_us")
cl <- "ifs_code"
FE <-NULL
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

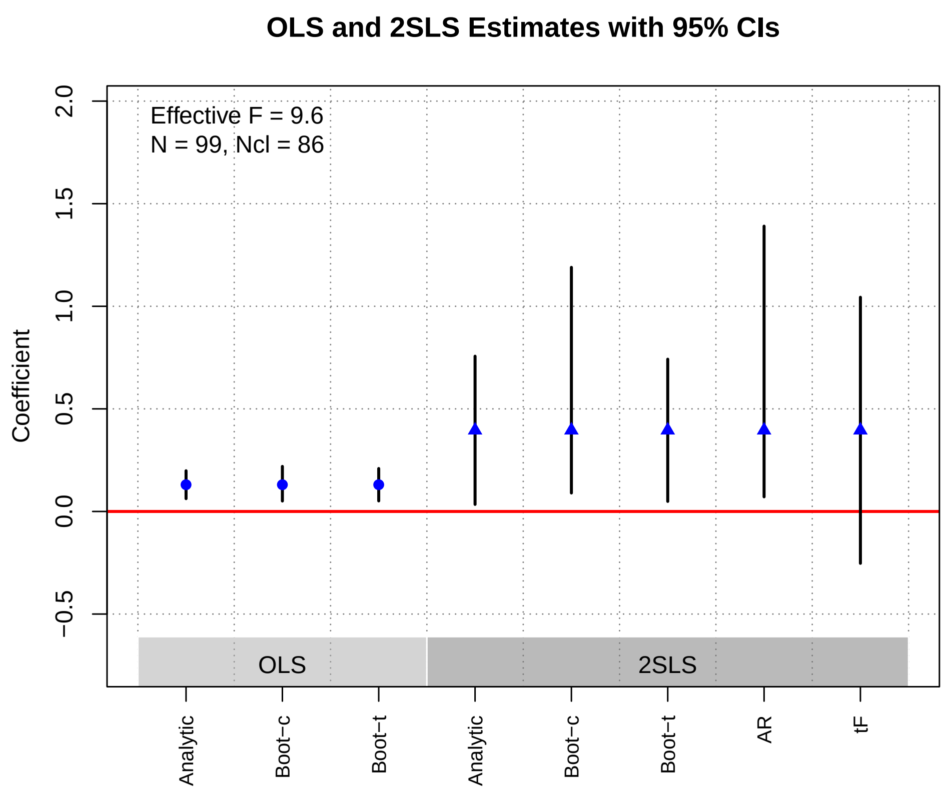
```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
```

```
## Bootstrap took 13.845 sec.
## AR Test Inversion...

## $est_ols
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.1304 0.0347 3.7620  0.0624   0.1983   2e-04
## Boot.c   0.1304 0.0410 3.1809  0.0509   0.2195   4e-03
## Boot.t   0.1304 0.0347 3.7620  0.0514   0.2093   1e-03
##
## $est_2sls
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.3956 0.1843 2.1468  0.0344   0.7567  0.0318
## Boot.c   0.3956 0.3102 1.2750  0.0900   1.1897  0.0040
## Boot.t   0.3956 0.1843 2.1468  0.0490   0.7421  0.0290
##
## $AR
## $AR$Fstat
##       F     df1     df2       p
##  5.4776  1.0000 97.0000  0.0213
##
## $AR$ci.print
## [1] "[0.0713, 1.3905]"
##
## $AR$ci
## [1] 0.07126795 1.39051194
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##      6.3713      9.2042      9.5714      8.4225      9.5714
##
## $rho
## [1] 0.2641
##
## $tF
##       F      cF    Coef      SE       t  CI2.5% CI97.5% p-value
##  9.5714  3.5187  0.3956  0.1843  2.1468 -0.2528  1.0439  0.2318
##
## $est_rf
##                       Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## military_first_alt -3.3385 1.4608  0.0223 1.4141  -6.5059   -0.7425     0.004
##
## $est_fs
##                       Coef    SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
```

```
## military_first_alt -8.4401 2.782  0.0024 2.9082 -14.2494   -2.9162         0
##
## $p_iv
## [1] 1
##
## $N
## [1] 99
##
## $N_cl
## [1] 86
##
## $df
## [1] 85
##
## $nvalues
##      gfcf_priv_gdp gov1_yrs military_first_alt
## [1,]            99       63                  2
```

```
plot_coef(g)
```



Grossman et al. (2017)

Replication Summary	
Unit of analysis	region * year
Treatment	government fragmentation
Instrument	the number of distinct landmasses; length of medium and small streams; over-time variation in the number of regional governments
Outcome	public goods provision
Model	Table1(8)

```
df<-readRDS("./data/jop_Grossman_2017.rds")
Y <- "ServicesCA"
D <- "ladminpc_l5"
Z <- c("lmeanMINUSi_adminpc_l6", "lmeanMINUSi_adminpc2_l6",
       "herf", "herf2", "llength", "llength2")
controls <- c("lpop_l", "wdi_urban_l", "lgdppc_l", "conflict_l",
              "dpi_state_l", "p_polity2_l",
              "loilpc_l", "aid_pc_l","al_ethnic")
cl <- "ccodecow"
FE <- "year"
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

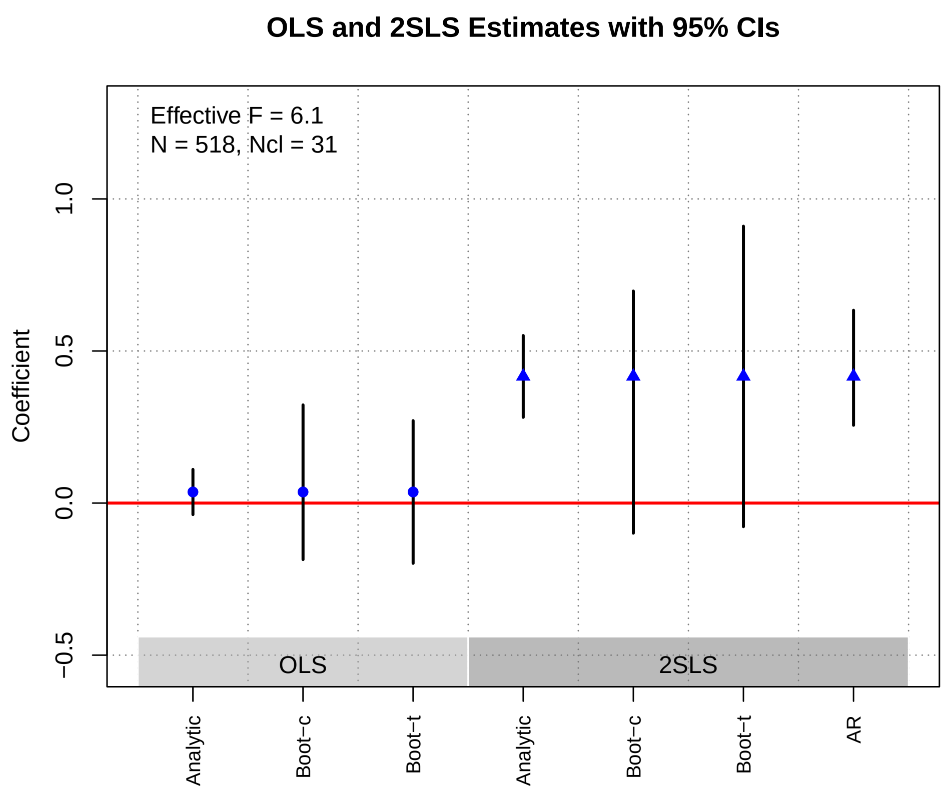
```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 29.803 sec.
## AR Test Inversion...

## $est_ols
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.0364 0.0379 0.9604 -0.0379   0.1107  0.3369
## Boot.c   0.0364 0.1309 0.2781 -0.1858   0.3228  0.7713
## Boot.t   0.0364 0.0379 0.9604 -0.1982   0.2710  0.7484
##
## $est_2sls
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.4164 0.0686 6.0671  0.2819   0.5509  0.0000
## Boot.c   0.4164 0.2050 2.0313 -0.0990   0.6972  0.1393
## Boot.t   0.4164 0.0686 6.0671 -0.0776   0.9104  0.1029
##
## $AR
## $AR$Fstat
##        F      df1      df2        p
##   9.4224   6.0000 511.0000   0.0000
##
## $AR$ci.print
```

```
## [1] "[0.2558, 0.6339]"
##
## $AR$ci
## [1] 0.2557837 0.6339076
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##     39.9978     40.9874     11.9593      1.0610      6.1390
##
## $rho
## [1] 0.581
##
## $est_rf
##                             Coef     SE p.value     SE.b  CI.b2.5% CI.b97.5%
## lmeanMINUSi_adminpc_l6   6.0801 6.5063  0.3500  11.4596  -19.3560   28.7435
## lmeanMINUSi_adminpc2_l6 -3.9097 1.9985  0.0504   3.1987  -10.4019    2.6593
## herf                    -0.0170 1.7102  0.9920 456.7395  -63.3790 1650.6832
## herf2                   -0.0545 1.2017  0.9638 236.1597 -853.3592   33.8186
## llength                  0.0669 0.0195  0.0006   0.8573   -0.9620    2.4337
## llength2                -0.0029 0.0014  0.0382   0.0320   -0.0932    0.0347
##                         p.value.b
## lmeanMINUSi_adminpc_l6     0.4636
## lmeanMINUSi_adminpc2_l6    0.1850
## herf                       0.6861
## herf2                      0.6341
## llength                    0.3909
## llength2                   0.4844
##
## $est_fs
##                              Coef     SE p.value     SE.b   CI.b2.5% CI.b97.5%
## lmeanMINUSi_adminpc_l6   27.1296 8.8691  0.0022  20.9404   -12.2898   72.9647
## lmeanMINUSi_adminpc2_l6 -13.3452 3.0243  0.0000   7.2207   -33.8902   -2.4266
## herf                      3.5973 2.4308  0.1389 386.3592 -1361.7025   70.1391
## herf2                    -2.4844 1.6773  0.1386 200.0866   -45.1780  696.6542
## llength                   0.0536 0.0162  0.0009   0.8918    -0.9538    2.4229
## llength2                  0.0002 0.0012  0.8936   0.0334    -0.0892    0.0398
##                         p.value.b
## lmeanMINUSi_adminpc_l6     0.1476
## lmeanMINUSi_adminpc2_l6    0.0125
## herf                       0.9709
## herf2                      0.9397
## llength                    0.5031
## llength2                   0.8503
```

```
##
## $p_iv
## [1] 6
##
## $N
## [1] 518
##
## $N_cl
## [1] 31
##
## $df
## [1] 476
##
## $nvalues
##      ServicesCA ladminpc_l5 lmeanMINUSi_adminpc_l6 lmeanMINUSi_adminpc2_l6 herf
## [1,]        518         518                    518                     518   15
##      herf2 llength llength2
## [1,]    15      29       29
```

```
plot_coef(g)
```



Healy and Malhotra (2013)

Replication Summary	
Unit of analysis	individual
Treatment	the share of a respondent's siblings who are female
Instrument	whether the younger sibling is a sister
Outcome	gender-role attitude in 1973
Model	Table1(1)

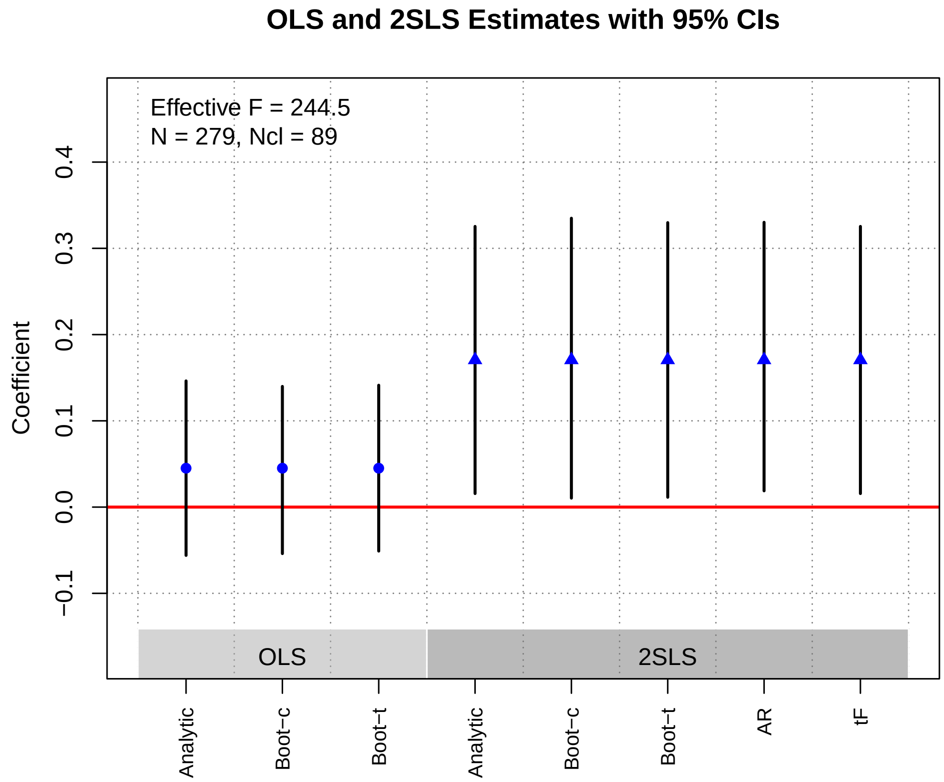
```
df <- readRDS("./data/jop_Healy_etal_2013.rds")
D <-"share_sis"
Y <- "womens_rights73"
Z <- "closest"
controls <- "num_sib"
cl <- "PSU"
FE <- NULL
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 12.956 sec.
## AR Test Inversion...
```

```
## $est_ols
##             Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.0451 0.0516 0.8748 -0.0560   0.1463  0.3817
## Boot.c   0.0451 0.0516 0.8747 -0.0539   0.1399  0.4060
## Boot.t   0.0451 0.0516 0.8748 -0.0510   0.1413  0.4030
##
## $est_2sls
##             Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.1706 0.0790 2.1589  0.0157   0.3254  0.0309
## Boot.c   0.1706 0.0821 2.0764  0.0104   0.3349  0.0400
## Boot.t   0.1706 0.0790 2.1589  0.0114   0.3298  0.0310
##
## $AR
## $AR$Fstat
##        F      df1      df2        p
##   4.8656   1.0000 277.0000   0.0282
##
## $AR$ci.print
## [1] "[0.0189, 0.3302]"
##
## $AR$ci
## [1] 0.01887136 0.33015989
##
```

```
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##    255.3329    252.1198    244.4704    237.9336    244.4704
##
## $rho
## [1] 0.6932
##
## $tF
##        F       cF     Coef       SE        t   CI2.5%  CI97.5%  p-value
## 244.4704   1.9600   0.1706   0.0790   2.1589   0.0157   0.3254   0.0309
##
## $est_rf
##           Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## closest 0.0832 0.0379  0.0281 0.0395   0.0051    0.1602      0.04
##
## $est_fs
##           Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## closest 0.4876 0.0307       0 0.0316   0.4255    0.5454         0
##
## $p_iv
## [1] 1
##
## $N
## [1] 279
##
## $N_cl
## [1] 89
##
## $df
## [1] 276
##
## $nvalues
##      womens_rights73 share_sis closest
## [1,]               7        17       2
```

```
plot_coef(g)
```



Henderson and Brooks (2016) (a)

Replication Summary	
Unit of analysis	district*year
Treatment	Democratic vote margins
Instrument	rain around election day
Outcome	incumbent roll call positioning
Model	Table3(1)

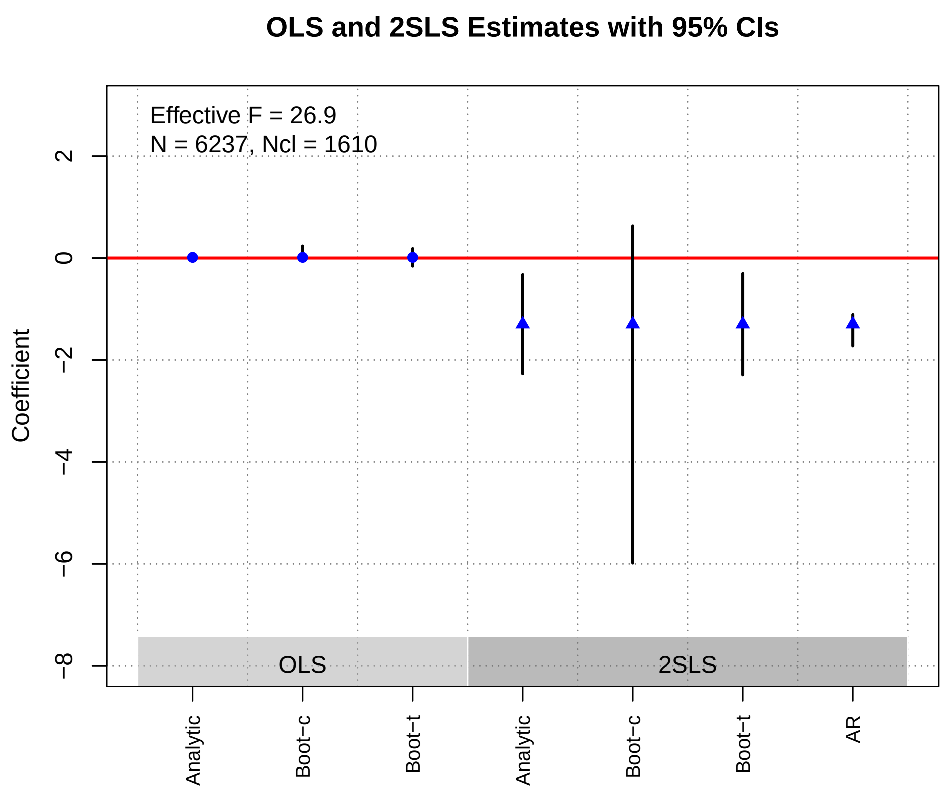
```
df<- readRDS("./data/jop_Henderson_etal_2016.rds")
df$fe_id_num<-df$`as.factor(fe_id_num)`
D <- "dose"
Y <- "vote"
Z <- c("rain_day", "rain_day_prev")
controls <- c("d_inc", "dist_prev", "midterm", "pres_party", "black",
              "construction", "educ", "minc", "farmer", "forborn",
              "gvtwkr", "manuf", "pop", "unempld", "urban", "retail",
              "sos", "gov", "comp_cq", "redistricted", "dose_prv", "vote_prv")
cl <- "fe_id_num" # incumbent
FE <- "fe_id_num"
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 44.196 sec.
## AR Test Inversion...
## Parallelising on 15 cores

## $est_ols
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.0124 0.0402 0.3089 -0.0664   0.0913  0.7574
## Boot.c   0.0124 0.0547 0.2274  0.0229   0.2354  0.0300
## Boot.t   0.0124 0.0402 0.3089 -0.1591   0.1839  0.9850
##
## $est_2sls
##             Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -1.2984 0.4963 -2.6159 -2.2712  -0.3255  0.0089
## Boot.c   -1.2984 2.2122 -0.5869 -5.9836   0.6294  0.1480
## Boot.t   -1.2984 0.4963 -2.6159 -2.2929  -0.3038  0.0210
##
## $AR
## $AR$Fstat
##         F       df1       df2         p
##   11.7276    2.0000 6234.0000    0.0000
##
## $AR$ci.print
## [1] "[-1.7252, -1.1097]"
##
## $AR$ci
## [1] -1.725204 -1.109747
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##     26.4294     21.5068     22.8295     11.2374     26.9117
##
## $rho
## [1] 0.1066
##
## $est_rf
##                  Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## rain_day      0.0326 0.0106  0.0021 0.0106   0.0179    0.0588     0.002
## rain_day_prev 0.0153 0.0089  0.0868 0.0121  -0.0251    0.0217     0.906
##
## $est_fs
##                   Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
```

```
## rain_day      -0.0144 0.0032       0 0.0043  -0.0194   -0.0028     0.008
## rain_day_prev -0.0187 0.0033       0 0.0045  -0.0191   -0.0019     0.014
##
## $p_iv
## [1] 2
##
## $N
## [1] 6237
##
## $N_cl
## [1] 1610
##
## $df
## [1] 1609
##
## $nvalues
##      vote dose rain_day rain_day_prev
## [1,] 6230 5138     5321          5326
```

```
plot_coef(g)
```



Henderson and Brooks (2016) (b)

Replication Summary	
Unit of analysis	district*year
Treatment	Democratic vote margins
Instrument	rain around election weekend
Outcome	incumbent roll call positioning
Model	Table3(2)

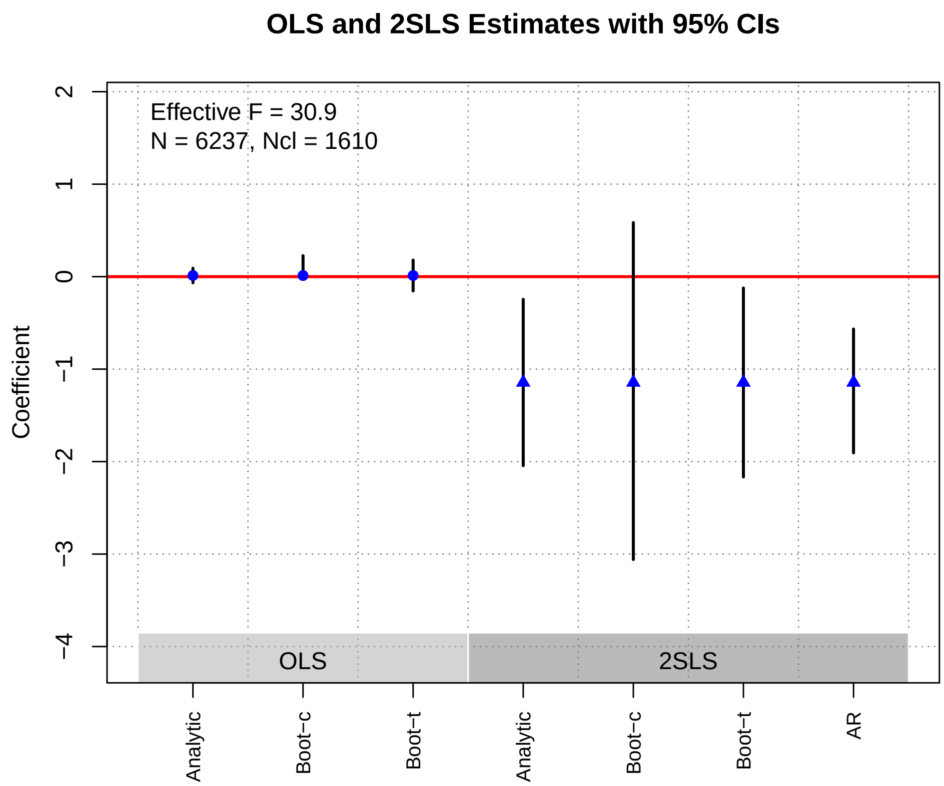
```r
df<- readRDS("./data/jop_Henderson_etal_2016.rds")
df$fe_id_num<-df$`as.factor(fe_id_num)`
D <- "dose"
Y <- "vote"
Z <- c("rain_weekend", "rain_weekend_prev")
controls <- c("d_inc", "dist_prev", "midterm", "pres_party", "black",
              "construction", "educ", "minc", "farmer", "forborn",
              "gvtwkr", "manuf", "pop", "unempld", "urban", "retail",
              "sos", "gov", "comp_cq", "redistricted", "dose_prv", "vote_prv")
cl <- "fe_id_num" # incumbent
FE <- "fe_id_num"
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 44.337 sec.
## AR Test Inversion...
## Parallelising on 15 cores

## $est_ols
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.0124 0.0402 0.3089 -0.0664   0.0913  0.7574
## Boot.c   0.0124 0.0519 0.2393  0.0241   0.2274  0.0180
## Boot.t   0.0124 0.0402 0.3089 -0.1549   0.1797  0.9770
##
## $est_2sls
##             Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic -1.1444 0.4588 -2.494 -2.0437  -0.2450  0.0126
## Boot.c   -1.1444 0.9395 -1.218 -3.0592   0.5846  0.1900
## Boot.t   -1.1444 0.4588 -2.494 -2.1661  -0.1226  0.0370
##
## $AR
## $AR$Fstat
##         F       df1       df2         p
##    8.6638    2.0000 6234.0000    0.0002
##
## $AR$ci.print
```

```
## [1] "[-1.9060, -0.5662]"
##
## $AR$ci
## [1] -1.9060260 -0.5662246
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##     30.3614     24.5741     26.3171     14.7219     30.9359
##
## $rho
## [1] 0.1141
##
## $est_rf
##                     Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## rain_weekend      0.0306 0.0117  0.0087 0.0115   0.0069    0.0516     0.010
## rain_weekend_prev 0.0175 0.0102  0.0867 0.0145  -0.0304    0.0253     0.866
##
## $est_fs
##                     Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## rain_weekend      -0.0192 0.0037       0 0.0047  -0.0254   -0.0070     0.000
## rain_weekend_prev -0.0213 0.0037       0 0.0046  -0.0232   -0.0053     0.004
##
## $p_iv
## [1] 2
##
## $N
## [1] 6237
##
## $N_cl
## [1] 1610
##
## $df
## [1] 1609
##
## $nvalues
##      vote dose rain_weekend rain_weekend_prev
## [1,] 6230 5138         5401              5407
```

```
plot_coef(g)
```



Johns and Pelc (2016)

Replication Summary	
Unit of analysis	WTO dispute
Treatment	the number third parties
Instrument	trade stake of the rest of the world
Outcome	becoming a third party
Model	Table2(2)

```
df<-readRDS("./data/jop_Johns_etal_2016.rds")
D='third_num_excl'
Y='thirdparty'
Z='ln_ROW_before_disp'
controls=c("ln_gdpk_partner", "ln_history_third", "ln_history_C",
    "Multilateral", "trade_before_dispute",  "ARTICLEXXII")
cl <- NULL
FE <- NULL
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

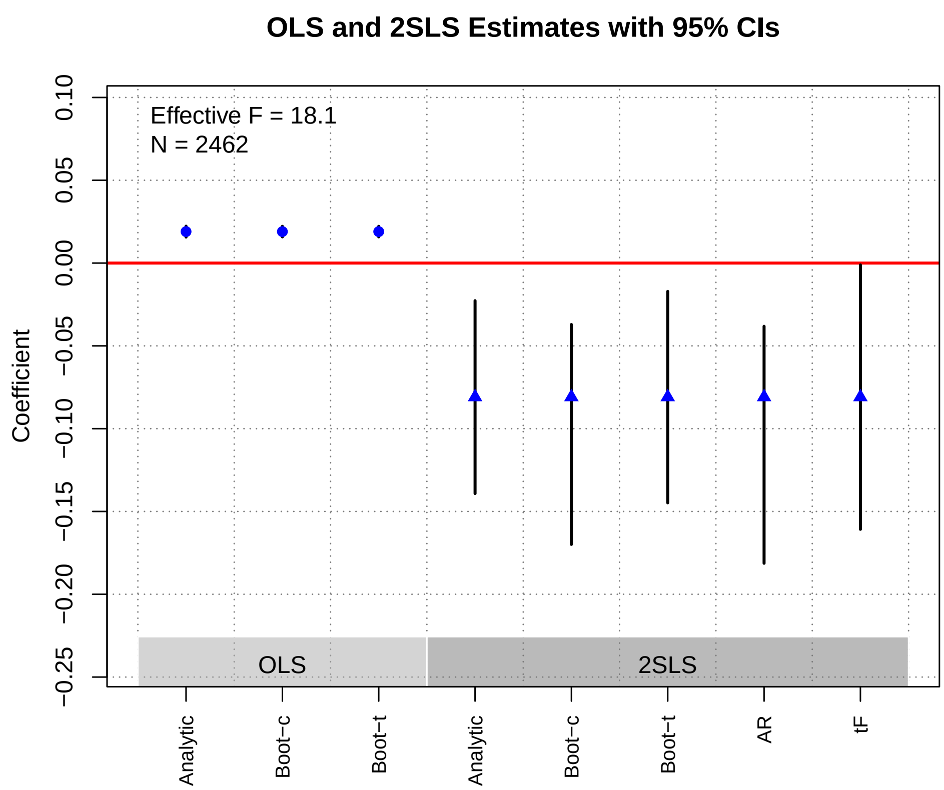
```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 13.166 sec.
```

```
## AR Test Inversion...

## $est_ols
##           Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic 0.019 0.0017 11.3469  0.0157   0.0223       0
## Boot.c   0.019 0.0016 11.5364  0.0158   0.0222       0
## Boot.t   0.019 0.0017 11.3469  0.0158   0.0222       0
##
## $est_2sls
##             Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -0.0809 0.0297 -2.7247 -0.1392  -0.0227  0.0064
## Boot.c   -0.0809 0.0338 -2.3952 -0.1699  -0.0371  0.0000
## Boot.t   -0.0809 0.0297 -2.7247 -0.1448  -0.0171  0.0200
##
## $AR
## $AR$Fstat
##         F       df1       df2         p
##   20.6337    1.0000 2460.0000    0.0000
##
## $AR$ci.print
## [1] "[-0.1813, -0.0382]"
##
## $AR$ci
## [1] -0.18130636 -0.03816286
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##     16.9224     18.1200          NA     19.5759     18.1200
##
## $rho
## [1] 0.0828
##
## $tF
##       F      cF    Coef      SE       t  CI2.5% CI97.5% p-value
## 18.1200  2.6873 -0.0809  0.0297 -2.7247 -0.1608 -0.0011  0.0469
##
## $est_rf
##                       Coef     SE p.value  SE.b CI.b2.5% CI.b97.5% p.value.b
## ln_ROW_before_disp -0.0137 0.0031       0 0.003  -0.0192   -0.0077         0
##
## $est_fs
##                      Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## ln_ROW_before_disp 0.1692 0.0397       0 0.0382   0.0976    0.2441         0
```

```
##
## $p_iv
## [1] 1
##
## $N
## [1] 2462
##
## $N_cl
## NULL
##
## $df
## [1] 2454
##
## $nvalues
##      thirdparty third_num_excl ln_ROW_before_disp
## [1,]          2             17               2281
```

```
plot_coef(g)
```



Kriner and Schickler (2014)

Replication Summary	
Unit of analysis	month
Treatment	committee investigations

Replication Summary	
Instrument	number of days that Congress was in session in a given month
Outcome	presidential approval
Model	Table1(1)

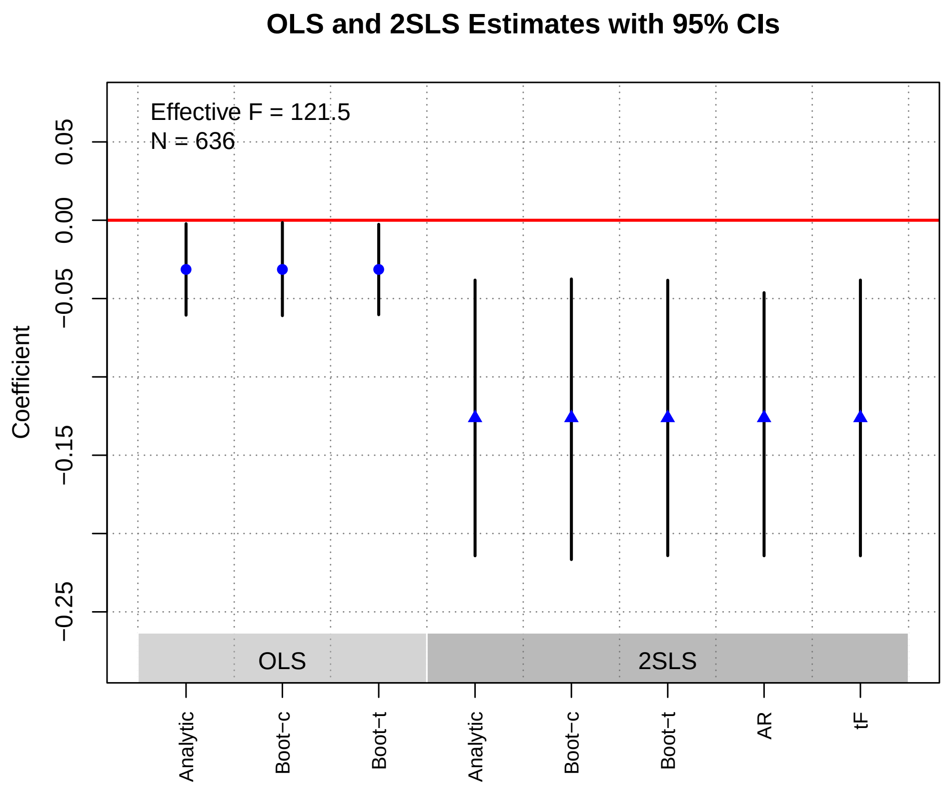
```r
df<-readRDS("./data/jop_Kriner_etal_2014.rds")
D <- "misconductdays"
Y <- "approval"
Z <- "alldaysinsession"
controls <- c("icst1", "positive", "negative", "vcaslast6mos",
              "iraqcaslast6mos", "honeymoon", "approvalt1", "ike","jfk",
              "lbj","rmn","ford","carter","reagan","bush","clinton","wbush")
cl <- NULL
FE <- NULL
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 13.627 sec.
## AR Test Inversion...
```

```
## $est_ols
##              Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -0.0314 0.0149 -2.1103 -0.0606  -0.0022  0.0348
## Boot.c   -0.0314 0.0149 -2.1056 -0.0609  -0.0015  0.0380
## Boot.t   -0.0314 0.0149 -2.1103 -0.0603  -0.0026  0.0320
##
## $est_2sls
##              Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -0.1262 0.0449 -2.8096 -0.2142  -0.0382   0.005
## Boot.c   -0.1262 0.0452 -2.7941 -0.2166  -0.0375   0.000
## Boot.t   -0.1262 0.0449 -2.8096 -0.2141  -0.0383   0.000
##
## $AR
## $AR$Fstat
##        F      df1      df2        p
##   9.6155   1.0000 634.0000   0.0020
##
## $AR$ci.print
## [1] "[-0.2142, -0.0462]"
##
## $AR$ci
## [1] -0.21418637 -0.04623729
##
```

```
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##    105.5872    121.5394          NA    117.2062    121.5394
##
## $rho
## [1] 0.382
##
## $tF
##        F       cF     Coef       SE        t   CI2.5%  CI97.5%  p-value
## 121.5394   1.9600  -0.1262   0.0449  -2.8096  -0.2142  -0.0382   0.0050
##
## $est_rf
##                    Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## alldaysinsession -0.035 0.0119  0.0032 0.0117  -0.0569    -0.011         0
##
## $est_fs
##                    Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## alldaysinsession 0.2777 0.0252       0 0.0257   0.2245     0.326         0
##
## $p_iv
## [1] 1
##
## $N
## [1] 636
##
## $N_cl
## NULL
##
## $df
## [1] 618
##
## $nvalues
##      approval misconductdays alldaysinsession
## [1,]      185             52               49
```

```
plot_coef(g)
```



Lei and Zhou (2022)

Replication Summary	
Unit of analysis	city*year
Treatment	subway approval
Instrument	whether the city has more than 3 million residents* population size
Outcome	mayor promotion
Model	Table3(A)

```
df<-readRDS("./data/jop_Lei_2022.rds")
Y <-'Mayor_promotion3y'
D <-'Mayor_plan'
Z <-'iv1'
controls<-c( 'Per_pop_2', 'iv1_int')
cl<-"City_Code"
FE<-c("provinceyear","City_Code")
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

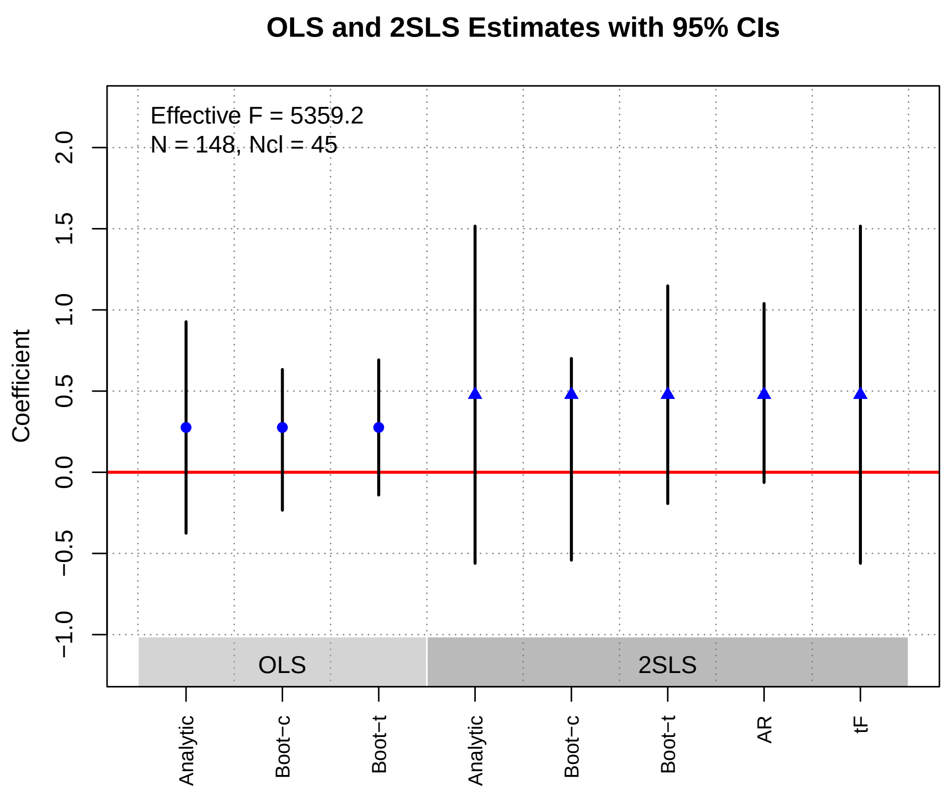
```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 24.208 sec.
```

```
## AR Test Inversion...

## $est_ols
##           Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.276 0.3323 0.8305 -0.3753   0.9273  0.4062
## Boot.c   0.276 0.2432 1.1348 -0.2337   0.6332  0.1505
## Boot.t   0.276 0.3323 0.8305 -0.1401   0.6921  0.1408
##
## $est_2sls
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.4776 0.5298 0.9014 -0.5608   1.5160  0.3674
## Boot.c   0.4776 0.3060 1.5608 -0.5413   0.7010  0.1602
## Boot.t   0.4776 0.5298 0.9014 -0.1927   1.1479  0.1044
##
## $AR
## $AR$Fstat
##        F      df1      df2        p
##   3.0197   1.0000 146.0000   0.0844
##
## $AR$ci.print
## [1] "[-0.0628, 1.0392]"
##
## $AR$ci
## [1] -0.06283252  1.03915838
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##     53.4747   2276.8055   5359.1714    179.8188   5359.1714
##
## $rho
## [1] 0.7604
##
## $tF
##         F        cF      Coef        SE         t    CI2.5%   CI97.5%   p-value
## 5359.1714    1.9600    0.4776    0.5298    0.9014   -0.5608    1.5160    0.3674
##
## $est_rf
##       Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## iv1 0.4833 0.5385  0.3695 0.3241  -0.6032    0.7417    0.1602
##
## $est_fs
##       Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## iv1 1.0119 0.0212       0 0.0755    0.992    1.2634         0
```

```
##
## $p_iv
## [1] 1
##
## $N
## [1] 148
##
## $N_cl
## [1] 45
##
## $df
## [1] 39
##
## $nvalues
##      Mayor_promotion3y Mayor_plan iv1
## [1,]                 2          2   2
```

```
plot_coef(g)
```



Lerman et al. (2017)

Replication Summary	
Unit of analysis	individual
Treatment	public versus only private health insurance

Replication Summary	
Instrument	born 1946 or 1947
Outcome	support ACA
Model	Table1(1)

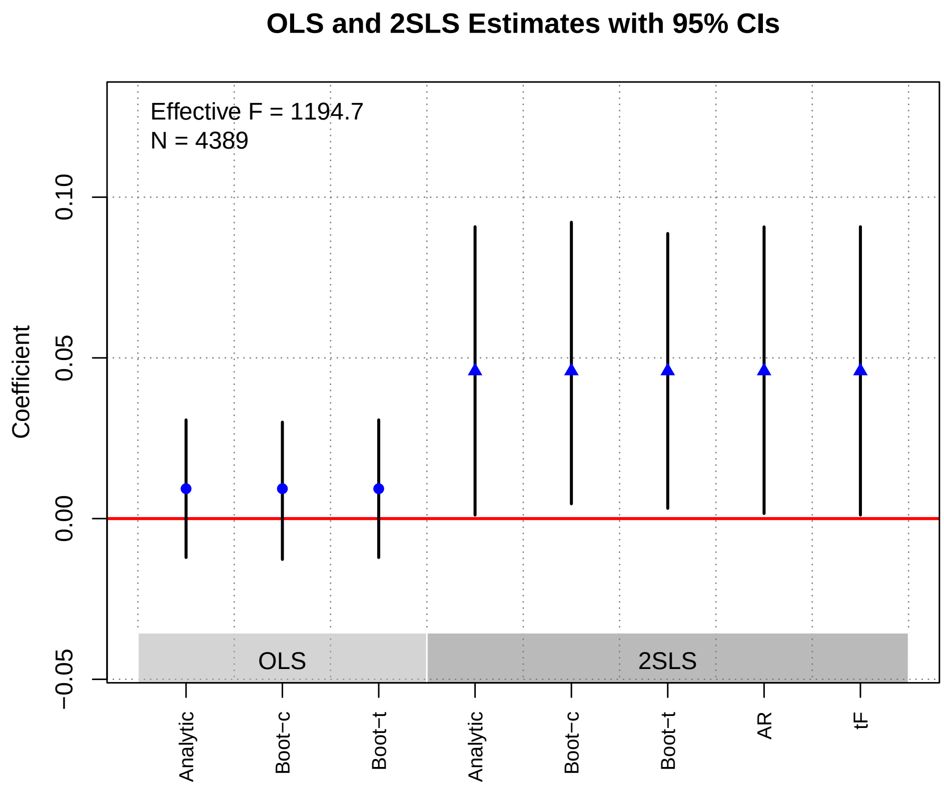
```
df<-readRDS("./data/jop_Lerman_2017.rds")
Y <-'suppafford'
D <-'privpubins3r'
Z <-'byr4647'
controls<-c( 'rep', 'ind', 'con', 'mod',
              'ideostrength', 'hcsocial', 'fininsur',
             'healthcaresupport', 'child18', 'male',
             'married', 'labor', 'mobility', 'homeowner',
             'religimp','employed', 'votereg', 'vote08',
             'black', 'hispanic2', 'military', 'educ',
              'fincome', 'newsint', 'publicemp', 'bornagain')
cl<-NULL
FE<-NULL
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 36.264 sec.
## AR Test Inversion...

## $est_ols
##              Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.0093 0.0109 0.8542 -0.0121   0.0307   0.393
## Boot.c   0.0093 0.0111 0.8368 -0.0127   0.0300   0.386
## Boot.t   0.0093 0.0109 0.8542 -0.0121   0.0307   0.393
##
## $est_2sls
##              Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.0459 0.0229 2.0095  0.0011   0.0908  0.0445
## Boot.c   0.0459 0.0225 2.0463  0.0046   0.0922  0.0280
## Boot.t   0.0459 0.0229 2.0095  0.0032   0.0887  0.0360
##
## $AR
## $AR$Fstat
##         F       df1       df2         p
##    4.0595    1.0000 4387.0000    0.0440
##
## $AR$ci.print
## [1] "[0.0016, 0.0908]"
```

```
##
## $AR$ci
## [1] 0.001589997 0.090751076
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##    1272.162    1194.659          NA    1122.025    1194.659
##
## $rho
## [1] 0.4752
##
## $tF
##         F        cF      Coef        SE         t    CI2.5%   CI97.5%   p-value
## 1194.6594    1.9600    0.0459    0.0229    2.0095    0.0011    0.0908    0.0445
##
## $est_rf
##           Coef   SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## byr4647 0.0202 0.01  0.0441 0.0098    0.002    0.0399     0.028
##
## $est_fs
##           Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## byr4647 0.4401 0.0127       0 0.0131   0.4135    0.4633         0
##
## $p_iv
## [1] 1
##
## $N
## [1] 4389
##
## $N_cl
## NULL
##
## $df
## [1] 4361
##
## $nvalues
##      suppafford privpubins3r byr4647
## [1,]          2            2       2
```

```
plot_coef(g)
```



Lorentzen et al. (2014)

Replication Summary	
Unit of analysis	city
Treatment	large firm dominance in 2007
Instrument	same variable measured in 1999
Outcome	pollution information transparency index
Model	Table1(2)

```
df<-readRDS("./data/jop_Lorentzen_2014.rds")
D <- "lfd2007"
Y <- "pitiave3"
Z <- "lfd99"
controls <- c("lbudgetrev", "lexpratio", "tertratio", "sat_air_pca")
cl <- NULL
FE <- NULL
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

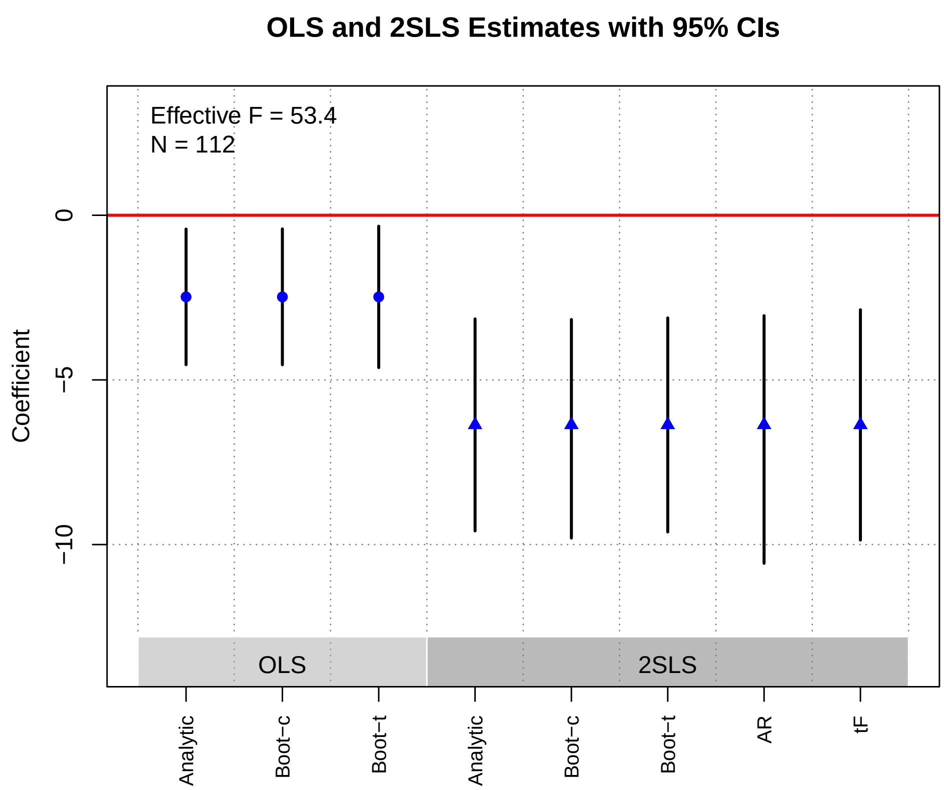
```
## Bootstrapping:
```

```
## Parallelising 1000 reps on 15 cores
## Bootstrap took 13.137 sec.
## AR Test Inversion...

## $est_ols
##             Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -2.4789 1.0508 -2.3590 -4.5385  -0.4193  0.0183
## Boot.c   -2.4789 1.0235 -2.4221 -4.5399  -0.4140  0.0180
## Boot.t   -2.4789 1.0508 -2.3590 -4.6259  -0.3318  0.0230
##
## $est_2sls
##             Coef     SE       t CI 2.5% CI 97.5% p.value
## Analytic -6.3664 1.6421 -3.8769 -9.5850  -3.1478   1e-04
## Boot.c   -6.3664 1.6918 -3.7632 -9.8039  -3.1658   0e+00
## Boot.t   -6.3664 1.6421 -3.8769 -9.6166  -3.1162   0e+00
##
## $AR
## $AR$Fstat
##        F      df1      df2        p
##  14.8495   1.0000 110.0000   0.0002
##
## $AR$ci.print
## [1] "[-10.5703, -3.0493]"
##
## $AR$ci
## [1] -10.570287  -3.049292
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##     53.6182     53.4100          NA     50.8991     53.4100
##
## $rho
## [1] 0.5796
##
## $tF
##       F      cF    Coef      SE       t  CI2.5% CI97.5% p-value
## 53.4100  2.1292 -6.3664  1.6421 -3.8769 -9.8628 -2.8700  0.0004
##
## $est_rf
##          Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## lfd99 -3.4227 0.8379       0 0.8356  -4.9772   -1.7695         0
##
## $est_fs
```

```
##          Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## lfd99 0.5376 0.0736       0 0.0754   0.3908    0.6837         0
##
## $p_iv
## [1] 1
##
## $N
## [1] 112
##
## $N_cl
## NULL
##
## $df
## [1] 106
##
## $nvalues
##      pitiave3 lfd2007 lfd99
## [1,]      108     112   112
```

```
plot_coef(g)
```



Pianzola et al. (2019)

Replication Summary	
Unit of analysis	individual
Treatment	smartvote use
Instrument	random assignment of the e-mail treatment
Outcome	vote intentions
Model	Table4(3)

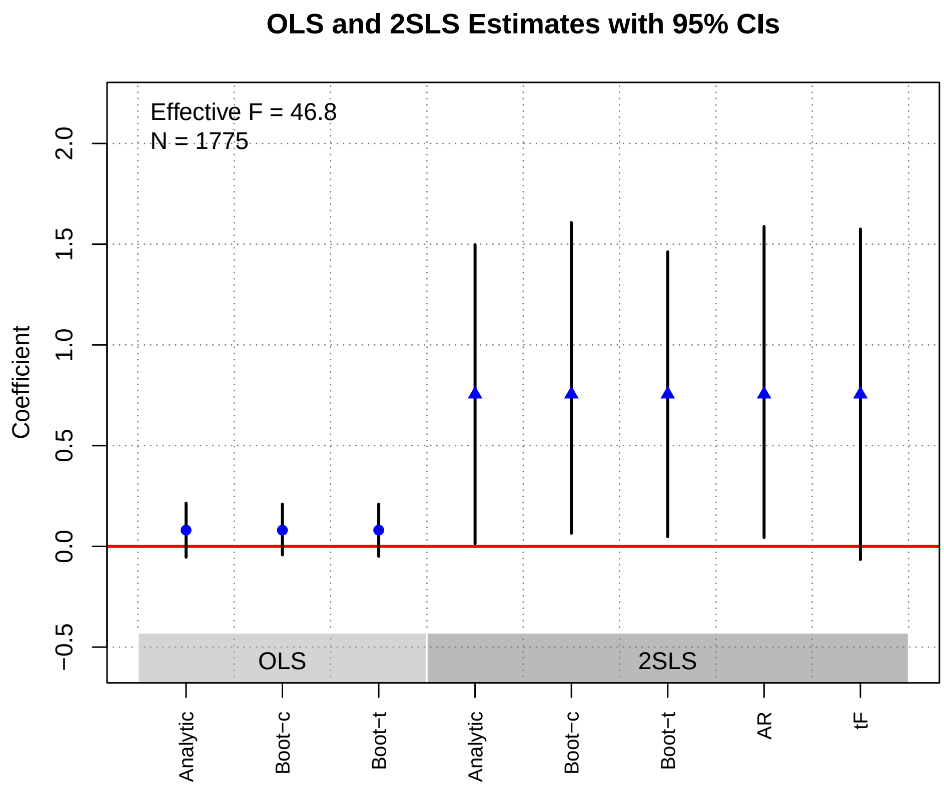
```
df <- readRDS("./data/jop_Pianzola_etal_2019.rds")
D <- "smartvote"
Y <- "diff_top_ptv"
Z <- "email"
controls <- NULL
cl <- NULL
FE <- NULL
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 13.606 sec.
## AR Test Inversion...

## $est_ols
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.0805 0.0684 1.1767 -0.0536   0.2146  0.2393
## Boot.c   0.0805 0.0659 1.2216 -0.0425   0.2100  0.2000
## Boot.t   0.0805 0.0684 1.1767 -0.0491   0.2101  0.2180
##
## $est_2sls
##           Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.755 0.3788 1.9934  0.0126   1.4974  0.0462
## Boot.c   0.755 0.3861 1.9555  0.0654   1.6072  0.0320
## Boot.t   0.755 0.3788 1.9934  0.0481   1.4620  0.0380
##
## $AR
## $AR$Fstat
##         F       df1       df2         p
##    4.2746    1.0000 1773.0000    0.0388
##
## $AR$ci.print
## [1] "[0.0429, 1.5883]"
##
## $AR$ci
## [1] 0.0429474 1.5883247
##
```

```
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##     46.7293     46.7612          NA     48.8745     46.7612
##
## $rho
## [1] 0.1602
##
## $tF
##       F      cF    Coef      SE       t  CI2.5% CI97.5% p-value
## 46.7612  2.1662  0.7550  0.3788  1.9934 -0.0654  1.5755  0.0713
##
## $est_rf
##         Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## email 0.1032 0.0499  0.0386 0.0492   0.0084    0.1996     0.032
##
## $est_fs
##         Coef   SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## email 0.1367 0.02       0 0.0195   0.0984    0.1754         0
##
## $p_iv
## [1] 1
##
## $N
## [1] 1775
##
## $N_cl
## NULL
##
## $df
## [1] 1773
##
## $nvalues
##      diff_top_ptv smartvote email
## [1,]           18         2     2
```

```
plot_coef(g)
```



Schleiter and Tavits (2016)

Replication Summary	
Unit of analysis	election
Treatment	opportunistic election calling
Instrument	prime Minister dissolution power
Outcome	vote share of Prime Minister's party
Model	Table3(b4)

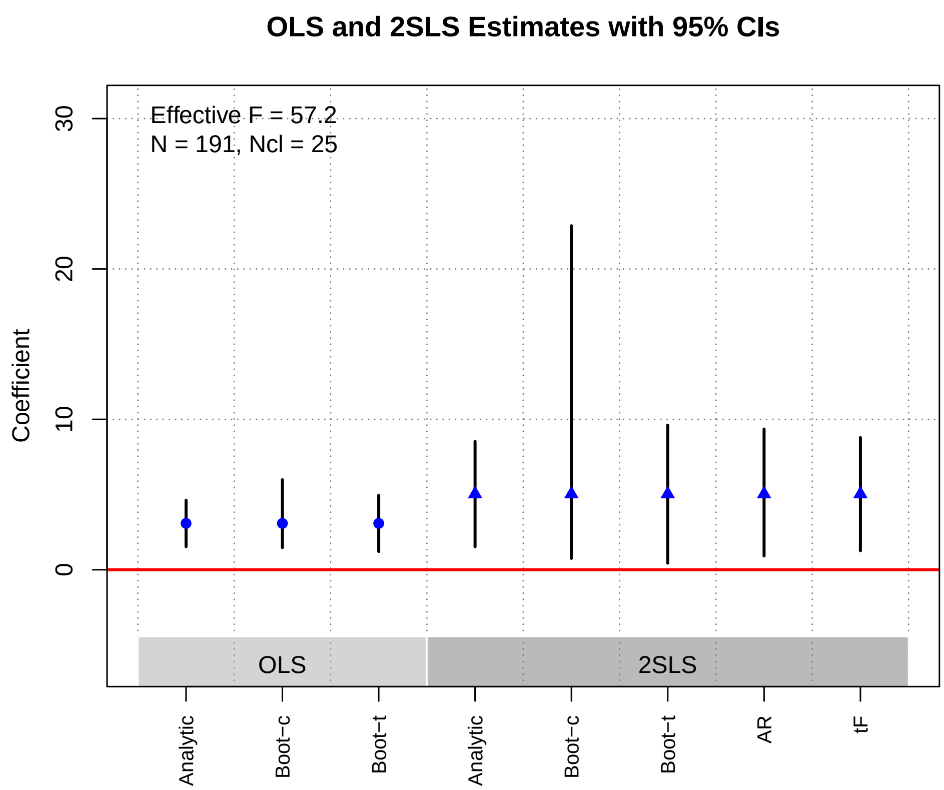
```
df<- readRDS("./data/jop_Schleiter_etal_2016.rds")
D <- "term2"
Y <- "pm_voteshare_next"
Z <- "disspm"
controls <- c("pm_voteshare", "gdp_chg1yr", "cpi1yr",  "dumcpi1yr")
cl <- "countryn"
FE <- "decade"
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 31.336 sec.
## AR Test Inversion...
```

```
## $est_ols
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 3.0828 0.7895 3.9048  1.5354   4.6302   1e-04
## Boot.c   3.0828 1.1217 2.7483  1.4746   5.9897   2e-03
## Boot.t   3.0828 0.7895 3.9048  1.2116   4.9540   2e-03
##
## $est_2sls
##            Coef      SE      t CI 2.5% CI 97.5% p.value
## Analytic 5.0282  1.7887 2.8111  1.5223   8.5341  0.0049
## Boot.c   5.0282 59.8193 0.0841  0.7576  22.8746  0.0320
## Boot.t   5.0282  1.7887 2.8111  0.4410   9.6155  0.0380
##
## $AR
## $AR$Fstat
##        F      df1      df2        p
##   5.7201   1.0000 189.0000   0.0178
##
## $AR$ci.print
## [1] "[0.9142, 9.3569]"
##
## $AR$ci
## [1] 0.9141871 9.3568882
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##    107.0322     75.6881     57.1949     21.4898     57.1949
##
## $rho
## [1] 0.6117
##
## $tF
##       F      cF    Coef      SE       t  CI2.5% CI97.5% p-value
## 57.1949  2.1037  5.0282  1.7887  2.8111  1.2653  8.7912  0.0088
##
## $est_rf
##          Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## disspm 0.3124 0.1062  0.0033 0.1807   0.0651     0.738     0.012
##
## $est_fs
##          Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## disspm 0.0621 0.0071       0 0.0134   0.0196    0.0745     0.024
##
## $p_iv
```

```
## [1] 1
##
## $N
## [1] 191
##
## $N_cl
## [1] 25
##
## $df
## [1] 179
##
## $nvalues
##      pm_voteshare_next term2 disspm
## [1,]               157     2      6
```

```
plot_coef(g)
```



Schubiger (2021)

Replication Summary	
Unit of analysis	community
Treatment	exposure to state violence
Instrument	location of a community inside or outside the emergency zone

Replication Summary	
Outcome	counterinsurgent mobilization

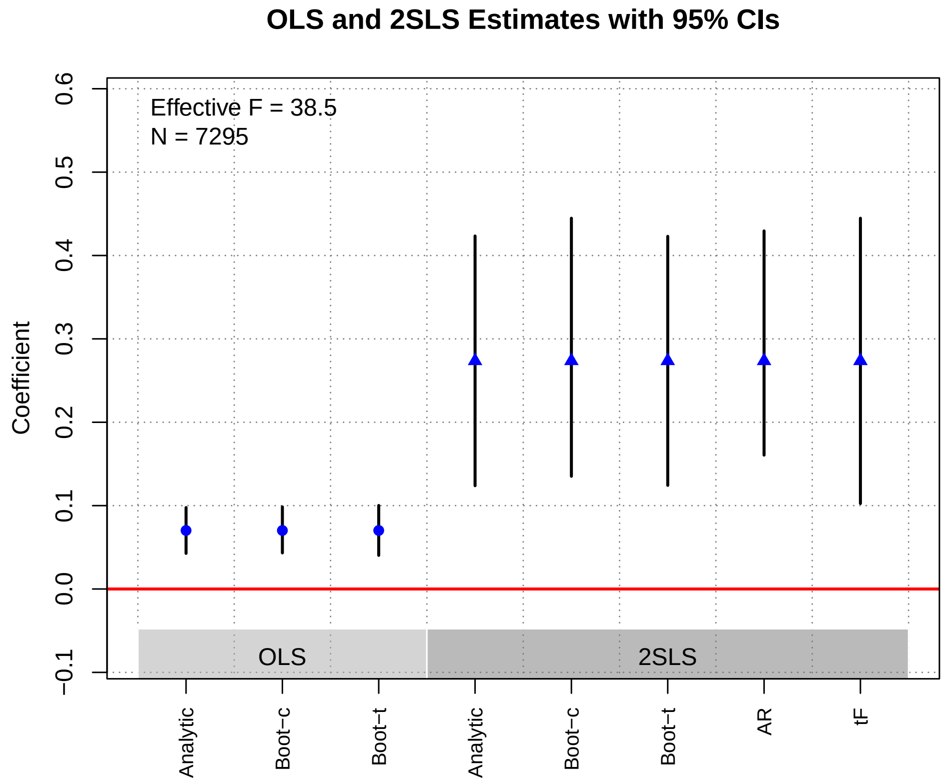
```
df <-readRDS("./data/jop_Schubiger_2021.rds")
D <- "violence_est_period2"
Y<-"autodefensa"
Z <- "emzone"
controls <-"distance"
cl<- NULL
FE<- NULL
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 13.601 sec.
## AR Test Inversion...
## Parallelising on 15 cores
```

```
## $est_ols
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.0702 0.0140 5.0069  0.0427   0.0977       0
## Boot.c   0.0702 0.0139 5.0638  0.0433   0.0985       0
## Boot.t   0.0702 0.0140 5.0069  0.0403   0.1002       0
##
## $est_2sls
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.2736 0.0764 3.5814  0.1239   0.4234   3e-04
## Boot.c   0.2736 0.0811 3.3745  0.1352   0.4447   0e+00
## Boot.t   0.2736 0.0764 3.5814  0.1243   0.4230   2e-03
##
## $AR
## $AR$Fstat
##         F       df1       df2         p
##   22.8597    1.0000 7293.0000    0.0000
##
## $AR$ci.print
## [1] "[0.1606, 0.4295]"
##
## $AR$ci
## [1] 0.1605585 0.4295009
##
## $AR$bounded
## [1] TRUE
##
```

```
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##     39.9899     38.5348          NA     40.6455     38.5348
##
## $rho
## [1] 0.0739
##
## $tF
##       F      cF    Coef      SE       t  CI2.5% CI97.5% p-value
## 38.5348  2.2392  0.2736  0.0764  3.5814  0.1025  0.4447  0.0017
##
## $est_rf
##          Coef     SE p.value  SE.b CI.b2.5% CI.b97.5% p.value.b
## emzone 0.0172 0.0048   4e-04 0.005    0.008    0.0273         0
##
## $est_fs
##          Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## emzone 0.0629 0.0101       0 0.0099   0.0428    0.0826         0
##
## $p_iv
## [1] 1
##
## $N
## [1] 7295
##
## $N_cl
## NULL
##
## $df
## [1] 7292
##
## $nvalues
##      autodefensa violence_est_period2 emzone
## [1,]           2                    2      2
```

```
plot_coef(g)
```



Stewart and Liou (2017)

Replication Summary	
Unit of analysis	insurgency*year
Treatment	foreign territory
Instrument	log total border length and the total number of that state's neighbors
Outcome	civilian casualties
Model	Table3(1)

```
df <- readRDS("./data/jop_Stewart_2017.rds")
D <- "exterrdum_low"
Y <- "oneside_best_log"
Z <- "total_border_ln"
controls <- c("bd_log", "terrdum", "strengthcent_ord", "rebstrength_ord",
              'nonmilsupport', 'rebestsize', 'l1popdensity',
              'l1gdppc_log','l1gdppc_change')
cl <- NULL
FE <- c("year","countrynum")
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

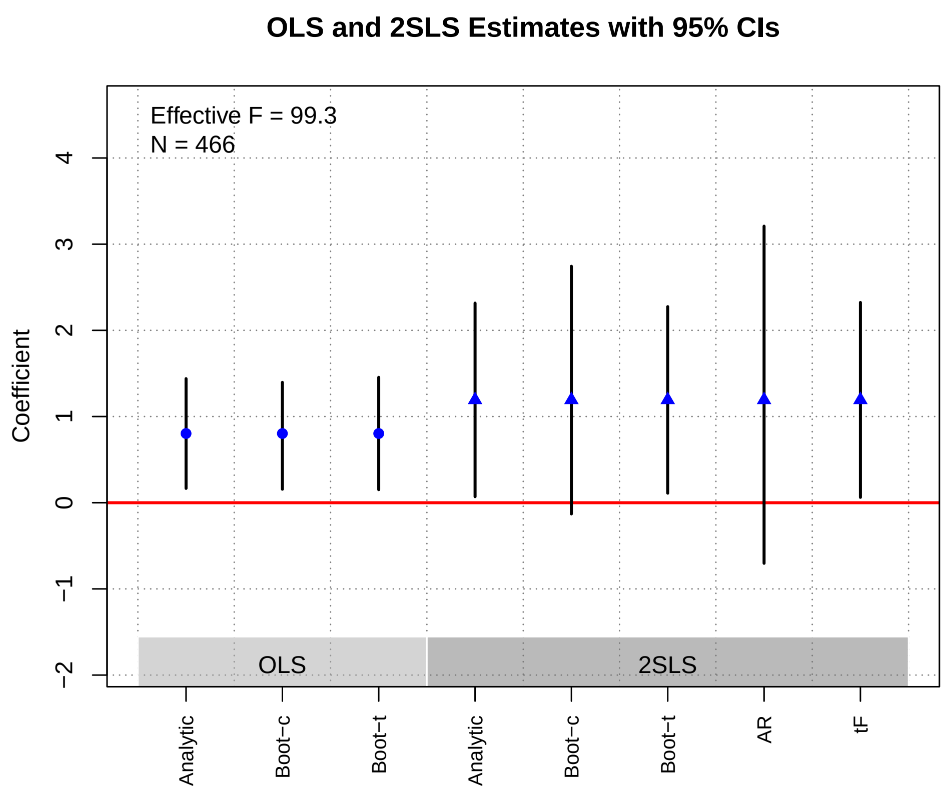
```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
```

```
## Bootstrap took 28.362 sec.
## AR Test Inversion...

## $est_ols
##           Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.803 0.3249 2.4716  0.1662   1.4398  0.0135
## Boot.c   0.803 0.3186 2.5201  0.1565   1.3963  0.0200
## Boot.t   0.803 0.3249 2.4716  0.1512   1.4548  0.0190
##
## $est_2sls
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 1.1929 0.5730 2.0817  0.0698   2.3161  0.0374
## Boot.c   1.1929 1.7071 0.6988 -0.1300   2.7437  0.0680
## Boot.t   1.1929 0.5730 2.0817  0.1108   2.2751  0.0290
##
## $AR
## $AR$Fstat
##       F     df1     df2       p
##   1.542   1.000 464.000   0.215
##
## $AR$ci.print
## [1] "[-0.7033, 3.2090]"
##
## $AR$ci
## [1] -0.7033116  3.2089531
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##     33.9859     99.3150          NA     63.5591     99.3150
##
## $rho
## [1] 0.2786
##
## $tF
##       F      cF    Coef      SE       t  CI2.5% CI97.5% p-value
## 99.3150  1.9734  1.1929  0.5730  2.0817  0.0621  2.3238  0.0387
##
## $est_rf
##                    Coef     SE p.value  SE.b CI.b2.5% CI.b97.5% p.value.b
## total_border_ln -7.0905 3.3952  0.0368 9.073 -15.5371    0.7844     0.068
##
## $est_fs
##                    Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
```

```
## total_border_ln -5.9438 0.5964       0 0.7456   -7.362   -4.6081         0
##
## $p_iv
## [1] 1
##
## $N
## [1] 466
##
## $N_cl
## NULL
##
## $df
## [1] 404
##
## $nvalues
##      oneside_best_log exterrdum_low total_border_ln
## [1,]              113             2              45
```

```
plot_coef(g)
```



Urpelainen and Zhang (2022)

Replication Summary	
Unit of analysis	district*year
Treatment	wind turbine capacity
Instrument	time trend multiplied by the wind resource of the electoral district
Outcome	Democratic vote
Model	Table3(B1)

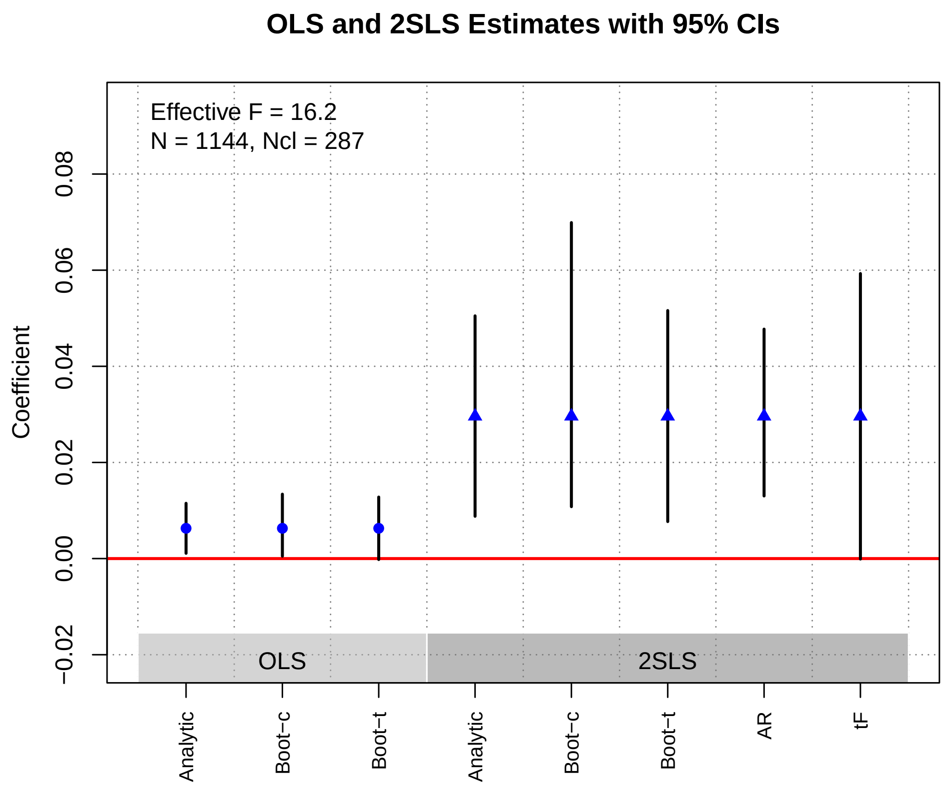
```r
df <-readRDS("./data/jop_urpelainen_2022.rds")
D <- "cum_capacity_turbine"
Y<-"demvotesmajorpercent"
Z <- "inter"
controls <-NULL
cl<- "district_fixed"
FE<- c("stateyear_fixed","district_fixed")
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 28.931 sec.
## AR Test Inversion...

## $est_ols
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.0063 0.0027 2.3711  0.0011   0.0115  0.0177
## Boot.c   0.0063 0.0034 1.8744  0.0005   0.0134  0.0380
## Boot.t   0.0063 0.0027 2.3711 -0.0002   0.0128  0.0620
##
## $est_2sls
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.0296 0.0106 2.7836  0.0088   0.0505  0.0054
## Boot.c   0.0296 0.0155 1.9108  0.0108   0.0699  0.0040
## Boot.t   0.0296 0.0106 2.7836  0.0077   0.0516  0.0080
##
## $AR
## $AR$Fstat
##         F       df1       df2         p
##   12.3395    1.0000 1142.0000    0.0005
##
## $AR$ci.print
## [1] "[0.0130, 0.0477]"
##
## $AR$ci
## [1] 0.01302764 0.04773662
```

```
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##     93.4366     27.8543     16.1654     15.4517     16.1654
##
## $rho
## [1] 0.3269
##
## $tF
##       F      cF    Coef      SE       t  CI2.5% CI97.5% p-value
## 16.1654  2.7897  0.0296  0.0106  2.7836 -0.0001  0.0593  0.0505
##
## $est_rf
##         Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## inter 0.9095 0.3122  0.0036 0.3238    0.295     1.563     0.004
##
## $est_fs
##          Coef     SE p.value  SE.b CI.b2.5% CI.b97.5% p.value.b
## inter 30.6883 5.8147       0 7.807  13.4343   44.1528         0
##
## $p_iv
## [1] 1
##
## $N
## [1] 1144
##
## $N_cl
## [1] 287
##
## $df
## [1] 286
##
## $nvalues
##      demvotesmajorpercent cum_capacity_turbine inter
## [1,]                  965                  141   777
```

```
plot_coef(g)
```



Webster et al. (2022)

Replication Summary	
Unit of analysis	individual
Treatment	percentage of angry words that a respondent wrote in his or her emotional recall prompt
Instrument	treatment assignment indicator
Outcome	social polarization: do favors
Model	Table2(1)

```
df <-readRDS("./data/jop_Webster_2022.rds")
D <- "anger"
Y<-"fourpack_1_01"
Z <- "treated"
controls <-"democrat"
cl<- NULL
FE<- NULL
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

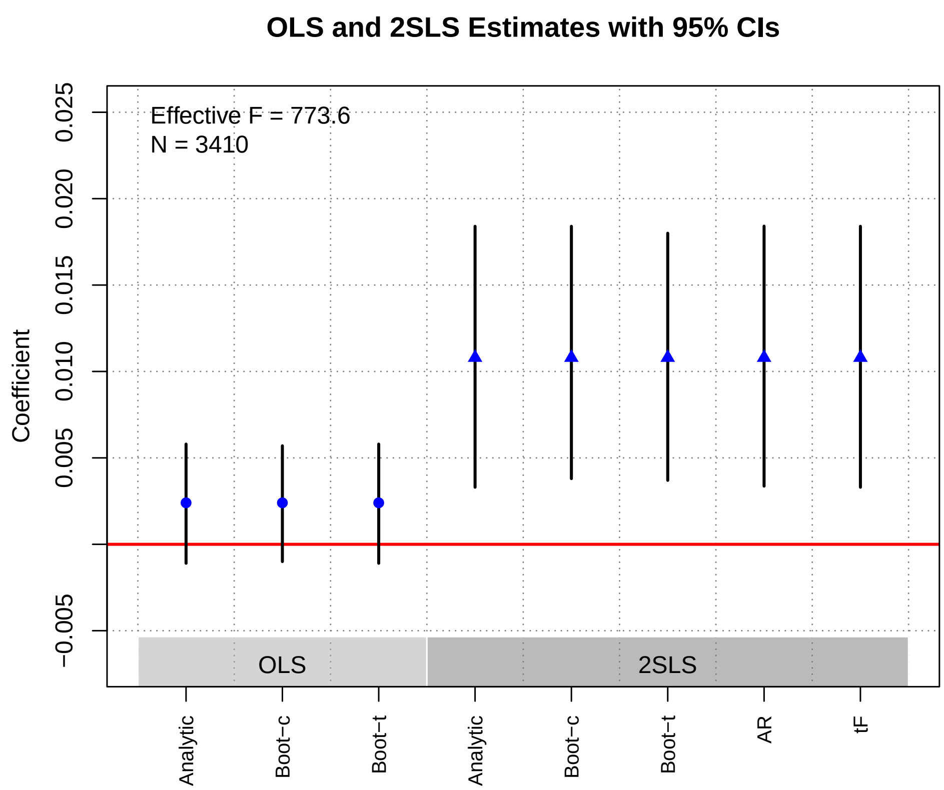
```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 12.876 sec.
```

```
## AR Test Inversion...

## $est_ols
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.0024 0.0018 1.3413 -0.0011   0.0058  0.1798
## Boot.c   0.0024 0.0017 1.3970 -0.0010   0.0057  0.1680
## Boot.t   0.0024 0.0018 1.3413 -0.0011   0.0058  0.1630
##
## $est_2sls
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.0108 0.0039 2.8123  0.0033   0.0184  0.0049
## Boot.c   0.0108 0.0037 2.9276  0.0038   0.0184  0.0020
## Boot.t   0.0108 0.0039 2.8123  0.0037   0.0180  0.0020
##
## $AR
## $AR$Fstat
##         F       df1       df2         p
##    8.0028    1.0000 3408.0000    0.0047
##
## $AR$ci.print
## [1] "[0.0034, 0.0184]"
##
## $AR$ci
## [1] 0.003365028 0.018409019
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##    801.9232    773.5894          NA    787.6258    773.5894
##
## $rho
## [1] 0.4365
##
## $tF
##        F       cF     Coef       SE        t   CI2.5%  CI97.5%  p-value
## 773.5894   1.9600   0.0108   0.0039   2.8123   0.0033   0.0184   0.0049
##
## $est_rf
##          Coef    SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## treated 0.031 0.011  0.0047 0.0105   0.0107     0.052     0.002
##
## $est_fs
##           Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## treated 2.8585 0.1028       0 0.1019    2.674     3.065         0
```

```
##
## $p_iv
## [1] 1
##
## $N
## [1] 3410
##
## $N_cl
## NULL
##
## $df
## [1] 3407
##
## $nvalues
##      fourpack_1_01 anger treated
## [1,]             5   252       2
```

```
plot_coef(g)
```



West (2017)

Replication Summary	
Unit of analysis	individual
Treatment	Obama win

Replication Summary	
Instrument	IEM (prediction market) price
Outcome	political efficacy
Model	Table1(4)

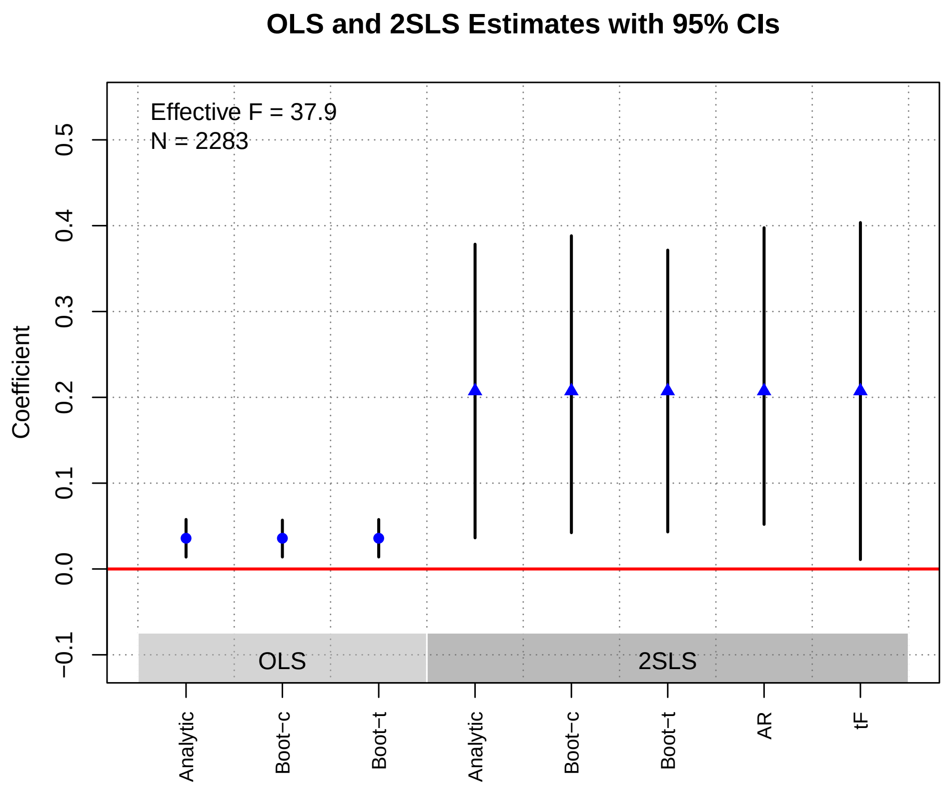
```
df<- readRDS("./data/jop_West_2017.rds")
D <- "obama"
Y <- "newindex"
Z <- "avgprice"
controls <- c("partyd1", "partyd2", "partyd3",
              "partyd4", "partyd5", "wa01_a", "wa02_a",
              "wa03_a", "wa04_a", "wa05_a",   "wfc02_a",
              "ra01_b",  "rd01", "wd02_b", "rkey",
              "wave_1", "dt_w12", "dt_w12_2")
cl <- NULL
FE <- c("state","religion")
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 29.262 sec.
## AR Test Inversion...

## $est_ols
##             Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.0358 0.0112 3.2084  0.0139   0.0577  0.0013
## Boot.c   0.0358 0.0112 3.1982  0.0140   0.0569  0.0000
## Boot.t   0.0358 0.0112 3.2084  0.0140   0.0576  0.0000
##
## $est_2sls
##             Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.2073 0.0873 2.3758  0.0363   0.3784  0.0175
## Boot.c   0.2073 0.0892 2.3234  0.0423   0.3882  0.0160
## Boot.t   0.2073 0.0873 2.3758  0.0432   0.3715  0.0180
##
## $AR
## $AR$Fstat
##         F       df1       df2         p
##    6.7445    1.0000 2281.0000    0.0095
##
## $AR$ci.print
## [1] "[0.0520, 0.3976]"
##
## $AR$ci
```

```
## [1] 0.05199789 0.39758798
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##     41.7917     37.8652          NA     39.5447     37.8652
##
## $rho
## [1] 0.1362
##
## $tF
##       F      cF    Coef      SE       t  CI2.5% CI97.5% p-value
## 37.8652  2.2493  0.2073  0.0873  2.3758  0.0110  0.4036  0.0384
##
## $est_rf
##            Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## avgprice 0.1407 0.0559  0.0119 0.0561   0.0273     0.246     0.016
##
## $est_fs
##            Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## avgprice 0.6784 0.1103       0 0.1079   0.4705    0.8891         0
##
## $p_iv
## [1] 1
##
## $N
## [1] 2283
##
## $N_cl
## NULL
##
## $df
## [1] 2211
##
## $nvalues
##      newindex obama avgprice
## [1,]      122     2      141
```

```
plot_coef(g)
```



Ziaja (2020)

Replication Summary	
Unit of analysis	country*year
Treatment	number of democracy donors
Instrument	constructed instrument
Outcome	democracy scores
Model	Table1(B2)

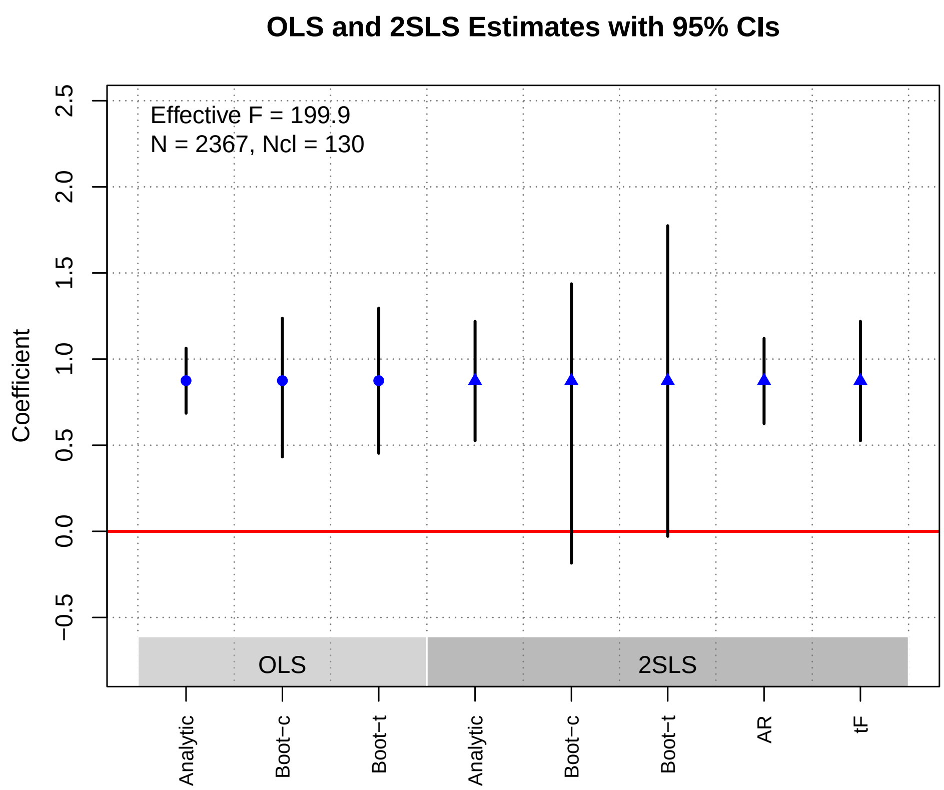
```
df <-readRDS("./data/jop_Ziaja_2020.rds")
D <- "l.CMgnh"
Y <- "v2x.polyarchy.n"
Z <- "l.ZwvCMgwh94"
controls <-c("l.pop.log.r", "l.gdpcap.log.r", "l.war25")
cl<- "cnamef"
FE<- c("cnamef", "periodf")
weights<-NULL
(g<-ivDiag(data=df, Y=Y, D=D, Z=Z, controls=controls, FE =FE,
  cl =cl,weights=weights, cores = cores))
```

```
## Bootstrapping:
## Parallelising 1000 reps on 15 cores
## Bootstrap took 30.623 sec.
## AR Test Inversion...
```

```
## $est_ols
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.8746 0.0963 9.0799  0.6858   1.0634   0.000
## Boot.c   0.8746 0.2027 4.3157  0.4320   1.2366   0.000
## Boot.t   0.8746 0.0963 9.0799  0.4532   1.2960   0.001
##
## $est_2sls
##            Coef     SE      t CI 2.5% CI 97.5% p.value
## Analytic 0.8726 0.1769 4.9338  0.5259   1.2192   0.000
## Boot.c   0.8726 0.4147 2.1043 -0.1846   1.4370   0.118
## Boot.t   0.8726 0.1769 4.9338 -0.0290   1.7742   0.059
##
## $AR
## $AR$Fstat
##         F       df1       df2         p
##   45.5711    1.0000 2365.0000    0.0000
##
## $AR$ci.print
## [1] "[0.6250, 1.1202]"
##
## $AR$ci
## [1] 0.6249779 1.1201789
##
## $AR$bounded
## [1] TRUE
##
##
## $F_stat
##  F.standard    F.robust   F.cluster F.bootstrap F.effective
##   1158.1467    775.0850    199.9166    208.9504    199.9166
##
## $rho
## [1] 0.586
##
## $tF
##        F       cF     Coef       SE        t   CI2.5%  CI97.5%  p-value
## 199.9166   1.9600   0.8726   0.1769   4.9338   0.5259   1.2192   0.0000
##
## $est_rf
##                 Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## l.ZwvCMgwh94 0.0599 0.0121       0 0.0301  -0.0121    0.1039     0.118
##
## $est_fs
##                 Coef     SE p.value   SE.b CI.b2.5% CI.b97.5% p.value.b
## l.ZwvCMgwh94 0.0686 0.0025       0 0.0047   0.0618    0.0804         0
##
## $p_iv
```

```
## [1] 1
##
## $N
## [1] 2367
##
## $N_cl
## [1] 130
##
## $df
## [1] 129
##
## $nvalues
##      v2x.polyarchy.n l.CMgnh l.ZwvCMgwh94
## [1,]            2038      24         2283
```

```
plot_coef(g)
```



References

Acharya, A., Blackwell, M., and Sen, M. (2016). The political legacy of american slavery. *The Journal of Politics*, 78(3):621–641. Publisher: University of Chicago Press Chicago, IL. Cited on pages 2 and 105.

Alt, J., Marshall, J., and Lassen, D. (2016). Credible sources and sophisticated voters: when does new information induce economic voting? *The Journal of Politics*, 78(2):327–342. Publisher: University of Chicago Press Chicago, IL. Cited on pages 2 and 108.

Arias, E. and Stasavage, D. (2019). How large are the political costs of fiscal austerity? *The Journal of Politics*, 81(4):1517–1522. Cited on pages 2 and 110.

Baccini, L. and Weymouth, S. (2021). Gone for good: Deindustrialization, white voter backlash, and us presidential voting. *American Political Science Review*, 115(2):550–567. Cited on pages 1 and 5.

Barth, E., Finseraas, H., and Moene, K. (2015). Political reinforcement: how rising inequality curbs manifested welfare generosity. *American Journal of Political Science*, 59(3):565–577. Publisher: Wiley Online Library. Cited on pages 1 and 42.

Bhavnani, R. and Lee, A. (2018). Local embeddedness and bureaucratic performance: evidence from india. *The Journal of Politics*, 80(1):71–87. Publisher: University of Chicago Press Chicago, IL. Cited on pages 2 and 113.

Blair, R. A., Di Salvatore, J., and Smidt, H. M. (2022). When do un peacekeeping operations implement their mandates? *American Journal of Political Science*, 66(3):664–680. Cited on pages 1 and 45.

Blattman, C., Hartman, A., and Blair, R. (2014). How to promote order and property rights under weak rule of law? an experiment in changing dispute resolution behavior through community education. *American Political Science Review*, page 100–120. Publisher: JSTOR. Cited on pages 1 and 7.

Carnegie, A. and Marinov, N. (2017). Foreign aid, human rights, and democracy promotion: Evidence from a natural experiment. *American Journal of Political Science*, 61(3):671–683. Publisher: Wiley Online Library. Cited on pages 1 and 47.

Charron, N., Dahlström, C., Fazekas, M., and Lapuente, V. (2017). Careers, connections, and corruption risks: Investigating the impact of bureaucratic meritocracy on public procurement processes. *The Journal of Politics*, 79(1):89–104. Publisher: University of Chicago Press Chicago, IL. Cited on pages 2 and 118.

Charron, N. and Lapuente, V. (2013). Why do some regions in europe have a higher quality of government? *The Journal of Politics*, 75(3):567–582. Publisher: Cambridge University Press New York, USA. Cited on pages 2 and 115.

Chong, A., León-Ciliotta, G., Roza, V., Valdivia, M., and Vega, G. (2019). Urbanization patterns, information diffusion, and female voting in rural paraguay. *American Journal of Political Science*, 63(2):323–341. Publisher: Wiley Online Library. Cited on pages 1 and 50.

Cirone, A. and Van Coppenolle, B. (2018). Cabinets, committees, and careers: the causal effect of committee service. *The Journal of Politics*, 80(3):948–963. Publisher: University of Chicago Press Chicago, IL. Cited on pages 2 and 120.

Colantone, I. and Stanig, P. (2018). Global competition and brexit. *American political science review*, 112(2):201–218. Cited on pages 1 and 10.

Colantone, I. and Stanig, P. (2018). The trade origins of economic nationalism: Import competition and voting behavior in western europe. *American Journal of Political Science*, 62(4):936–953. Publisher: Wiley Online Library. Cited on pages 1 and 53.

Coppock, A. and Green, D. (2016). Is voting habit forming? new evidence from experiments and regression discontinuities. *American Journal of Political Science*, 60(4):1044–1062. Publisher: Wiley Online Library. Cited on pages 1 and 55.

Croke, K., Grossman, G., Larreguy, H., and Marshall, J. (2016). Deliberate disengagement: How education can decrease political participation in electoral authoritarian regimes. *American Political Science Review*, 110(3):579–600. Publisher: Cambridge University Press. Cited on pages 1 and 12.

De La O, A. (2013). Do conditional cash transfers affect electoral behavior? evidence from a randomized experiment in mexico. *American Journal of Political Science*, 57(1):1–14. Publisher: Wiley Online Library. Cited on pages 1 and 58.

Dietrich, S. and Wright, J. (2015). Foreign aid allocation tactics and democratic change in africa. *The Journal of Politics*, 77(1):216–234. Publisher: University of Chicago Press Chicago, IL. Cited on pages 2 and 123.

DiGiuseppe, M. and Shea, P. E. (2022). Us patronage, state capacity, and civil conflict. *The Journal of Politics*, 84(2):767–782. Cited on pages 2 and 126.

Dower, P., Finkel, E., Gehlbach, S., and Nafziger, S. (2018). Collective action and representation in autocracies: Evidence from russia's great reforms. *American Political Science Review*, 112(1):125–147. Publisher: Cambridge University Press. Cited on pages 1, 15 and 17.

Dube, O. and Naidu, S. (2015). Bases, bullets, and ballots: The effect of us military aid on political conflict in colombia. *The Journal of Politics*, 77(1):249–267. Publisher: University of Chicago Press Chicago, IL. Cited on pages 2 and 128.

Feigenbaum, J. and Hall, A. (2015). How legislators respond to localized economic shocks: Evidence from chinese import competition. *The Journal of Politics*, 77(4):1012–1030. Publisher: University of Chicago Press Chicago, IL. Cited on pages 2 and 131.

Flores-Macias, G. and Kreps, S. (2013). The foreign policy consequences of trade: China's commercial relations with africa and latin america, 1992–2006. *The Journal of Politics*, 75(2):357–371. Publisher: Cambridge University Press New York, USA. Cited on pages 2 and 133.

Gehlbach, S. and Keefer, P. (2012). Private investment and the institutionalization of collective action in autocracies: ruling parties and legislatures. *The Journal of Politics*, 74(2):621–635. Publisher: Cambridge University Press New York, USA. Cited on pages 2 and 136.

Gerber, A., Huber, G., and Washington, E. (2010). Party affiliation, partisanship, and political beliefs: A field experiment. *American Political Science Review*, 104(4):720–744. Publisher: Cambridge University Press. Cited on pages 1 and 20.

Goldstein, R. and You, H. (2017). Cities as lobbyists. *American Journal of Political Science*, 61(4):864–876. Publisher: Wiley Online Library. Cited on pages 1 and 60.

Grossman, G., Pierskalla, J., and Boswell Dean, E. (2017). Government fragmentation and public goods provision. *The Journal of Politics*, 79(3):823–840. Publisher: University of Chicago Press Chicago, IL. Cited on pages 2 and 138.

Hager, A. and Hilbig, H. (2019). Do inheritance customs affect political and social inequality? *American Journal of Political Science*, 63(4):758–773. Publisher: Wiley Online Library. Cited on pages 2, 63 and 65.

Hager, A. and Krakowski, K. (2022). Does state repression spark protests? evidence from secret police surveillance in communist poland. *American Political Science Review*, 116(2):564–579. Cited on pages 1 and 25.

Hager, A., Krakowski, K., and Schaub, M. (2019). Ethnic riots and prosocial behavior: Evidence from kyrgyzstan. *American Political Science Review*, 113(4):1029–1044. Cited on pages 1 and 22.

Healy, A. and Malhotra, N. (2013). Childhood socialization and political attitudes: Evidence from a natural experiment. *The Journal of Politics*, 75(4):1023–1037. Publisher: Cambridge University Press New York, USA. Cited on pages 2 and 141.

Henderson, J. and Brooks, J. (2016). Mediating the electoral connection: The information effects of voter signals on legislative behavior. *The Journal of Politics*, 78(3):653–669. Cited on pages 2, 144 and 146.

Hong, J. Y., Park, S., and Yang, H. (2022). In strongman we trust: The political legacy of the new village movement in south korea. *American Journal of Political Science*. Cited on pages 2 and 68.

Johns, L. and Pelc, K. (2016). Fear of crowds in world trade organization disputes: Why don't more countries participate? *The Journal of Politics*, 78(1):88–104. Publisher: University of Chicago Press Chicago, IL. Cited on pages 2 and 149.

Kapoor, S. and Magesan, A. (2018). Independent candidates and political representation in india. *American Political Science Review*, 112(3):678–697. Publisher: Cambridge University Press. Cited on pages 1 and 27.

Kim, J. (2019). Direct democracy and women's political engagement. *American Journal of Political Science*, 63(3):594–610. Publisher: Wiley Online Library. Cited on pages 2 and 70.

Kocher, M., Pepinsky, T., and Kalyvas, S. (2011). Aerial bombing and counterinsurgency in the vietnam war. *American Journal of Political Science*, 55(2):201–218. Publisher: Wiley Online Library. Cited on pages 2 and 73.

Kriner, D. and Schickler, E. (2014). Investigating the president: Committee probes and presidential approval, 1953–2006. *The Journal of Politics*, 76(2):521–534. Publisher: Cambridge University Press New York, USA. Cited on pages 2 and 151.

Kuipers, N. and Sahn, A. (2022). The representational consequences of municipal civil service reform. *American Political Science Review*, pages 1–17. Cited on pages 1 and 30.

Laitin, D. and Ramachandran, R. (2016). Language policy and human development. *American Political Science Review*, 110(3):457–480. Publisher: Cambridge University Press. Cited on pages 1 and 32.

Lei, Z. and Zhou, J. A. (2022). Private returns to public investment: Political career incentives and infrastructure investment in china. *The Journal of Politics*, 84(1):455–469. Cited on pages 2 and 154.

Lelkes, Y., Sood, G., and Iyengar, S. (2017). The hostile audience: The effect of access to broadband internet on partisan affect. *American Journal of Political Science*, 61(1):5–20. Publisher: Wiley Online Library. Cited on pages 2 and 75.

Lerman, A., Sadin, M., and Trachtman, S. (2017). Policy uptake as political behavior: evidence from the affordable care act. *The American Political Science Review*, 111(4):755. Publisher: Cambridge University Press. Cited on pages 2 and 156.

López-Moctezuma, G., Wantchekon, L., Rubenson, D., Fujiwara, T., and Pe Lero, C. (2020). Policy deliberation and voter persuasion: Experimental evidence from an election in the philippines. *American Journal of Political Science*. Cited on pages 2 and 78.

Lorentzen, P., Landry, P., and Yasuda, J. (2014). Undermining authoritarian innovation: the power of china's industrial giants. *The Journal of Politics*, 76(1):182–194. Publisher: Cambridge University Press New York, USA. Cited on pages 2 and 159.

McClendon, G. (2014). Social esteem and participation in contentious politics: A field experiment at an lgbt pride rally. *American Journal of Political Science*, 58(2):279–290. Publisher: Wiley Online Library. Cited on pages 2 and 80.

Meredith, M. (2013). Exploiting friends-and-neighbors to estimate coattail effects. *American Political Science Review*, page 742–765. Publisher: JSTOR. Cited on pages 1 and 35.

Nellis, G. and Siddiqui, N. (2018). Secular party rule and religious violence in pakistan. *The American Political Science Review*, 112(1):49. Publisher: Cambridge University Press. Cited on pages 1 and 37.

Pianzola, J., Trechsel, A., Vassil, K., Schwerdt, G., and Alvarez, R. (2019). The impact of personalized information on vote intention: Evidence from a randomized field experiment. *The Journal of Politics*, 81(3):833–847. Publisher: The University of Chicago Press Chicago, IL. Cited on pages 2 and 161.

Ritter, E. and Conrad, C. (2016). Preventing and responding to dissent: The observational challenges of explaining strategic repression. *American Political Science Review*, 110(1):85–99. Publisher: Cambridge University Press. Cited on pages 1 and 40.

Rueda, M. (2017). Small aggregates, big manipulation: Vote buying enforcement and collective monitoring. *American Journal of Political Science*, 61(1):163–177. Publisher: Wiley Online Library. Cited on pages 2 and 83.

Schleiter, P. and Tavits, M. (2016). The electoral benefits of opportunistic election timing. *The Journal of Politics*, 78(3):836–850. Publisher: University of Chicago Press Chicago, IL. Cited on pages 2 and 164.

Schubiger, L. I. (2021). State violence and wartime civilian agency: Evidence from peru. *The Journal of Politics*, 83(4):1383–1398. Cited on pages 2 and 166.

Sexton, R., Wellhausen, R., and Findley, M. (2019). How government reactions to violence worsen social welfare: evidence from peru. *American Journal of Political Science*, 63(2):353–367. Publisher: Wiley Online Library. Cited on pages 2 and 85.

Spenkuch, J. and Tillmann, P. (2018). Elite influence? religion and the electoral success of the nazis. *American Journal of Political Science*, 62(1):19–36. Publisher: Wiley Online Library. Cited on pages 2 and 88.

Stewart, M. and Liou, Y. (2017). Do good borders make good rebels? territorial control and civilian casualties. *The Journal of Politics*, 79(1):284–301. Publisher: University of Chicago Press Chicago, IL. Cited on pages 2 and 169.

Stokes, L. (2016). Electoral backlash against climate policy: A natural experiment on retrospective voting and local resistance to public policy. *American Journal of Political Science*, 60(4):958–974. Publisher: Wiley Online Library. Cited on pages 2 and 90.

Tajima, Y. (2013). The institutional basis of intercommunal order: Evidence from indonesia's democratic transition. *American Journal of Political Science*, 57(1):104–119. Publisher: Wiley Online Library. Cited on pages 2 and 93.

Trounstine, J. (2016). Segregation and inequality in public goods. *American Journal of Political Science*, 60(3):709–725. Publisher: Wiley Online Library. Cited on pages 2 and 96.

Urpelainen, J. and Zhang, A. T. (2022). Electoral backlash or positive reinforcement? wind power and congressional elections in the united states. *The Journal of Politics*, 84(3):1306–1321. Cited on pages 2 and 171.

Vernby, K. (2013). Inclusion and public policy: Evidence from sweden's introduction of noncitizen suffrage. *American Journal of Political Science*, 57(1):15–29. Publisher: Wiley Online Library. Cited on pages 2 and 98.

Webster, S. W., Connors, E. C., and Sinclair, B. (2022). The social consequences of political anger. *The Journal of Politics*, 84(3):1292–1305. Cited on pages 2 and 174.

West, E. (2017). Descriptive representation and political efficacy: Evidence from obama and clinton. *The Journal of Politics*, 79(1):351–355. Publisher: University of Chicago Press Chicago, IL. Cited on pages 3 and 176.

Wood, A. K. and Grose, C. R. (2022). Campaign finance transparency affects legislators' election outcomes and behavior. *American Journal of Political Science*, 66(2):516–534. Cited on pages 2 and 100.

Zhu, B. (2017). Mncs, rents, and corruption: Evidence from china. *American Journal of Political Science*, 61(1):84–99. Publisher: Wiley Online Library. Cited on pages 2 and 103.

Ziaja, S. (2020). More donors, more democracy. *The Journal of Politics*, 82(2):433–447. Publisher: The University of Chicago Press Chicago, IL. Cited on pages 3 and 179.