Online Supporting Information

Achieving Statistical Significance with Control Variables and without Transparency

Gabriel Lenz
Alexander Sahn

June 01, 2020

Contents

Quotes on Suppression Effects **3**
Quotes on Controls Introducing Bias 3

Method **5**
Article Selection Rules 5
Model Selection Rules 5

Table S1: Reasons for Exclusion **6**

Figure S1: P-value Changes in All Studies **7**

Figure S2: P-value Decreases from Coefficient Change Only **8**

Figure S3: Control Effects in Bivariate Versus Full Specifications **9**

Figure S4: Changes in Coefficients across All Studies **10**

Figure S5: Changes in Coefficients that Do Not Show Bivariate **11**

Figure S6: T-Value Changes in Observational Studies **12**

Figure S7a: Confidence Intervals of Bivariate and Full Specifications **13**

Figure S7b: Confidence Intervals of Bivariate and Full Specifications (Normed Absolute Value Coefficients) **14**

Figure S7c: Confidence Intervals of Bivariate and Full Specifications (Absolute Value Beta Coefficients) **15**

Figure S7d: Confidence Intervals of Bivariate and Full Specifications (Absolute Value Beta Coefficients–Zoomed) **16**

Figure S8: Multicollinearity in Observational Studies **17**

Specification Uncertainty **18**
Statistically Significant in All Specifications 18
Standard Deviation of Estimates 19

Analyses Specified in Pre-Analysis Plan **20**
Inflating with Controls 20
Sensitivity to Control Choice 24
Deflators and Inflators 24

Articles' Text about Controls **25**

Quotes on Suppression Effects

"Ideally, including suppressor variables in a model should be theory based and every regression model should include using a test for suppressor effects. This approach allows researchers to become ware of the suppressor effect of a particular variable and to be better able to explain when regression results change drastically from one model to another." (Pandey and Elliott, 2010)

"Suppression and multicollinearity present challenging conditions for those who wish to apply multiple regression analysis." (Beckstead, 2012)

"The authors discuss definitions of the suppressor phenomenon, show how the unwary researcher can be warned against it, and present guidelines for the interpretation of the results." (Maassen and Bakker, 2001)

"In his influential book, Personality and Prediction, Jerry Wiggins (1973) concluded:"the case for suppressor variables remains to be demonstrated" (p. 38). The impact of the book on a generation of personality researchers led them, quite rightly, to be skeptical about the value of suppressor variables. Wiggins's conclusion was based partly on Ghiselli's (1972) failure to replicate a suppressor effect, leading him to liken suppressor variables to the ephemeral 'will-o-the wisp' (p. 270). Wiggins was particularly persuaded by the disappointing results reported by Goldberg, Rorer, and Green (1970). Since those earlier warnings, summary evaluations of the utility of suppressor variables in behavioral science have remained guarded at best (e.g., Cohen & Cohen, 1992; Pedhazur, 1982). Even very recently, Maassen and Baker (2001) warned against devoting energy to formulating theoretical explanations for solitary suppression results." (Paulhus et al., 2004)

"The most difficult interpretive problem, however, occurs in the face of suppression when relations between an independent variable and a dependent variable increase or change direction following partialling (e.g., Cohen & Cohen, 1983; Darlington, 1968; Horst, 1941). This situation poses the greatest interpretive hazard because a relation that did not exist or did not exist as strongly before partialling is now uncovered" (Lynam, Hoyle and Newman, 2006).

"Suppression effects in multiple linear regression are one of the most elusive phenomena in the educational and psychological measurement literature." (Kim, 2019)

We note that some of these authors are using suppression effects to describe a very particular case where the suppressor variable is uncorrelated with the dependent variable.

Quotes on Controls Introducing Bias

"Omitting both the fixed effects and the unobserved confounder was preferable to adjusting for the fixed effects precisely because the two biases counterbalanced one another in the unadjusted estimate. In practice, a researcher is unlikely to know whether adjusting for covariates will unmask unobserved confounder bias. Similar observations have led to

somewhat pessimistic assessment of observational analysis, for example, in Clarke (2005) and Frisell et al. (2012) (but see also Clarke [2009]).” (Middleton et al., 2016)

“Using a simple linear (regression) setting with two confounders one observed (X), the other unobserved (U) we demonstrate that conditioning on the observed confounder X does not necessarily imply that the confounding bias decreases, even if X is highly correlated with U. That is, adjusting for X may increase instead of reduce the omitted variable bias (OVB). Two phenomena can cause an increasing OVB: (i) bias amplification and (ii) cancellation of offsetting biases.” (Steiner and Kim, 2016)

“A key underlying assumption is that the danger posed by omitted variable bias can be ameliorated by the inclusion of relevant control variables. Unfortunately, as this article demonstrates, there is nothing in the mathematics of regression analysis that supports this conclusion. The inclusion of additional control variables may increase or decrease the bias, and we cannot know for sure which is the case in any particular situation.” (Clarke, 2005)

“Scholars often assume that the danger posed by omitted variable bias can be ameliorated by the inclusion of large numbers of relevant control variables. However, there is nothing in the mathematics of regression analysis that supports this conclusion. This paper goes beyond textbook treatments of omitted variable bias and shows, both for OLS and for generalized linear models, that the inclusion of additional control variables may increase or decrease the bias, and we cannot know for sure which is the case in any particular situation. The last section of the paper shows how formal sensitivity analysis can be used to determine whether omitted variables are a problem. A substantive example demonstrates the method.” (Clarke, 2009)

Method

According to our pre-analysis plan (see anonymous link: https://goo.gl/tCeNFM), we selected articles and models from within these articles according to the following rules. We mark aspects of our plan that evolved as we collected the data in italics.

Article Selection Rules

We included *AJPS* articles from volumes 57-59, because the strict replication data and code posting was fully in effect for publish articles by volume 57.

We selected articles according to the following rules:

- The article must rely on a statistical model (excludes experiments without controls, formal models, political theory, simulations, etc.).
- The article's analysis uses *a standard statistical model* and replication code and data is available.
- The article's analysis uses three or more controls.
  - *To increase the number of studies, we lowered the threshold to just one control.*
- The article's publication rest primary on a single, non-null result (excludes findings driven by multiple results or null results).

Table S1 presents a count of the articles excluded by the reason for exclusion.

Model Selection Rules

Given that papers often have multiple hypotheses and multiple tests of each hypothesis, we selected one statistical model from each paper according to the following criteria:

- We selected a model that is the main finding of the paper-referenced in the abstract or in body.
- If there are multiple specifications of the same model with different sets of controls, we chose the model with the full set of controls.
- If there are multiple cases or measures of the key independent variable and the text does not point to one specifically, *we used the summary measure, if the paper presents one, and if not,* we used the model with the smallest absolute value coefficient on the key variable.
- We included *any variables necessary for the research design or identification strategy*.

Table S1: Reasons for Exclusion

Data or code unavailable	6
Formal model	19
Measurement focus	3
No control variables	35
Nonstandard statistical model	13
Null finding	2
Plethora of claims	14
Simulation	2
Theory	5
Total articles excluded	99

"Nonstandard statistical model" are those without a function in Stata or R.

"Plethora of claims" describes articles that did not depend on one key statistically significant result, but on several results.

We exclude all AJPS "workshop" articles.

Figure S1: P-value Changes in All Studies

In the plot below, we show the average change in p-values from the bivariate to the full specification for *all* articles, not just those that do not disclose the bivariate. The figure first shows p-value changes all studies, observational and experimental, with 95% confidence intervals. It next shows the key finding in the paper, that observational studies showing a bivariate specification have only a trivial amount of p-value decreases, but that observational studies that do not show a bivariate specification have a larger amount. Next, it shows that the p-value decreases occur in studies with fixed effects and without fixed effects. It then shows the p-value changes for all three *AJPS* volumes we analyze and all three subfields. Finally, it shows estimates with below the median number of controls and above the median number of controls. The mean number of controls in these studies is nine and the median is eight. In the paper, we show similar statistics only for studies that did not to show a bivariate specification.

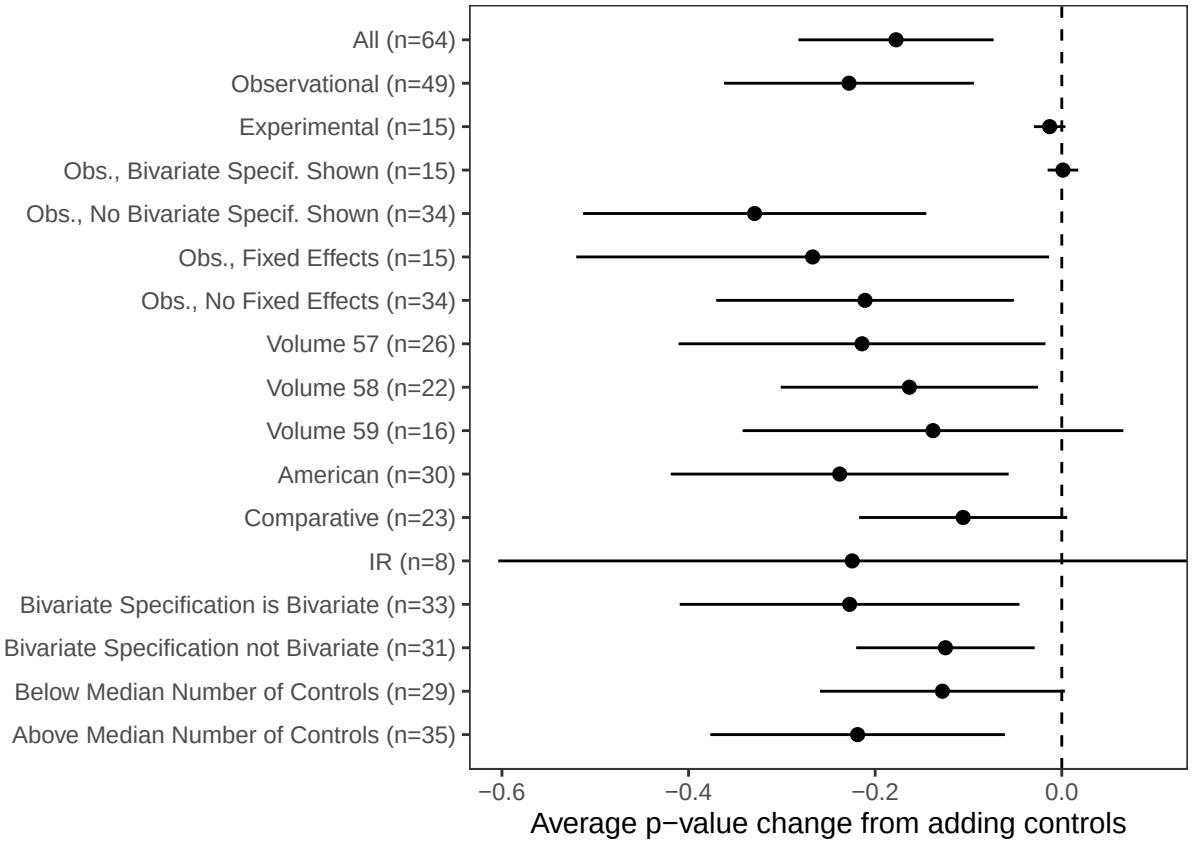


Figure S2: P-value Decreases from Coefficient Change Only

In the plot below, we show the average change in p-values from the bivariate to the full specification (with 95% confidence intervals) for observational and experimental studies due only to coefficient change. The figure shows that the majority of the p-value change shown in the figures above are due to increases in the magnitude of the coefficient rather than a reduction in the standard errors. This finding is robust across different subsets of studies.

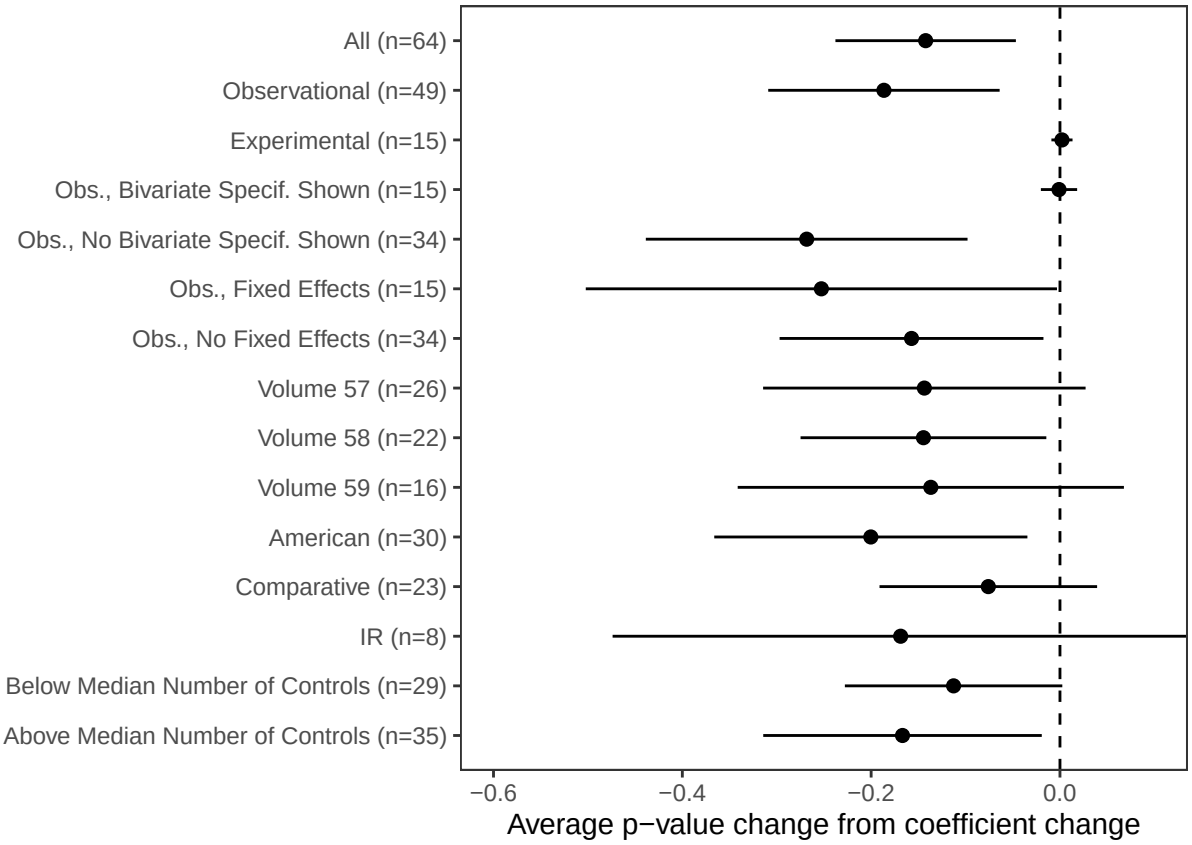


Figure S3: Control Effects in Bivariate Versus Full Specifications

As we discuss in the paper, one potential objection to our argument is that researchers can't justify/explain suppression effects because of the complexity of the multivariate space. This figure shows otherwise. It reveals that controls generally have similar effects on the key coefficient estimate in the simple bivariate case and in the complex multivariate case. Each point represents a control in a regression model, with the x value representing the effect of removing that control from the full specification and the y value representing the effect of adding it to the bivariate specification. All effects are standardized to a mean of zero and standard deviation of one. The dotted line shows the 45 degree line. The figure shows a strong relationship between these two, implying that explanations about the effect of controls on key effect estimates often translate from the bivariate to the full specification.

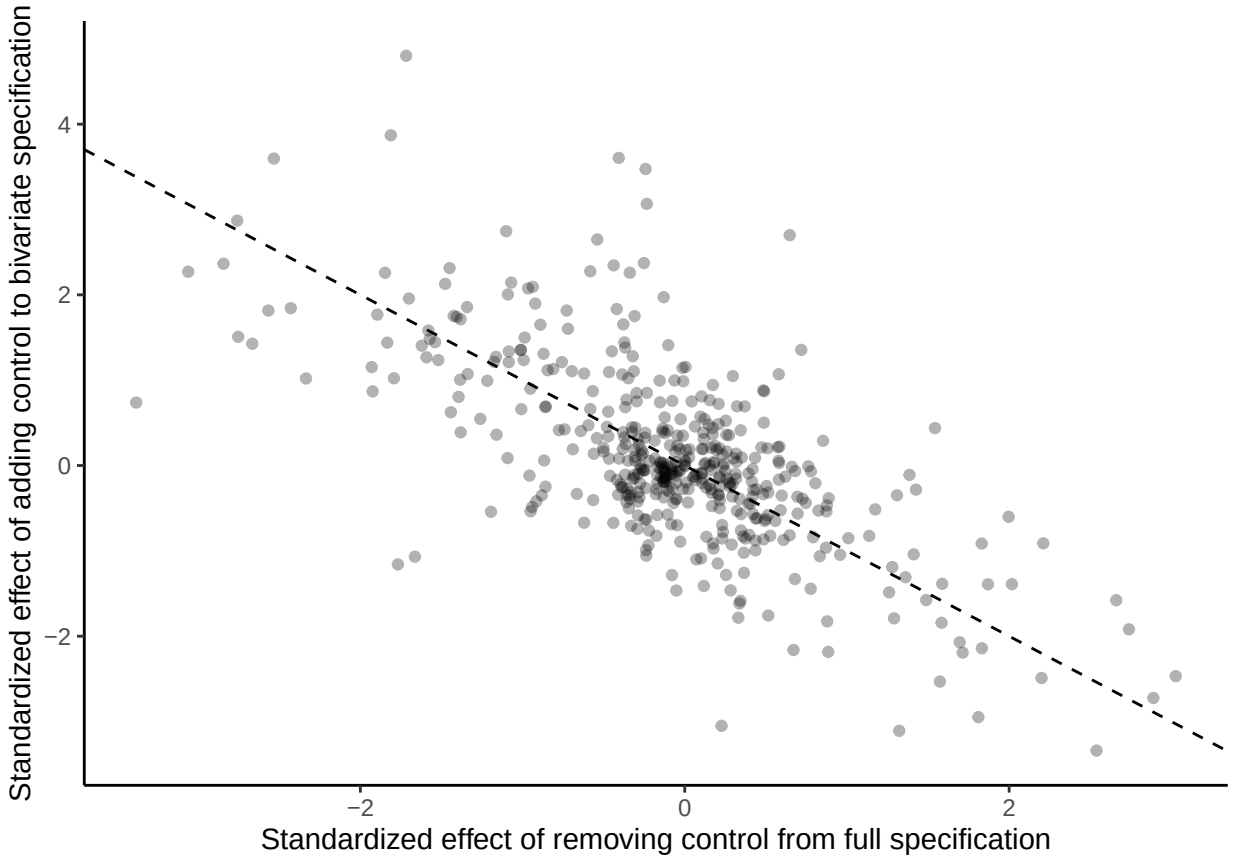


Figure S4: Changes in Coefficients across All Studies

This figure is analogous to S1 but for coefficient changes instead of p-value changes. We calculate coefficient percent changes with log plus one.

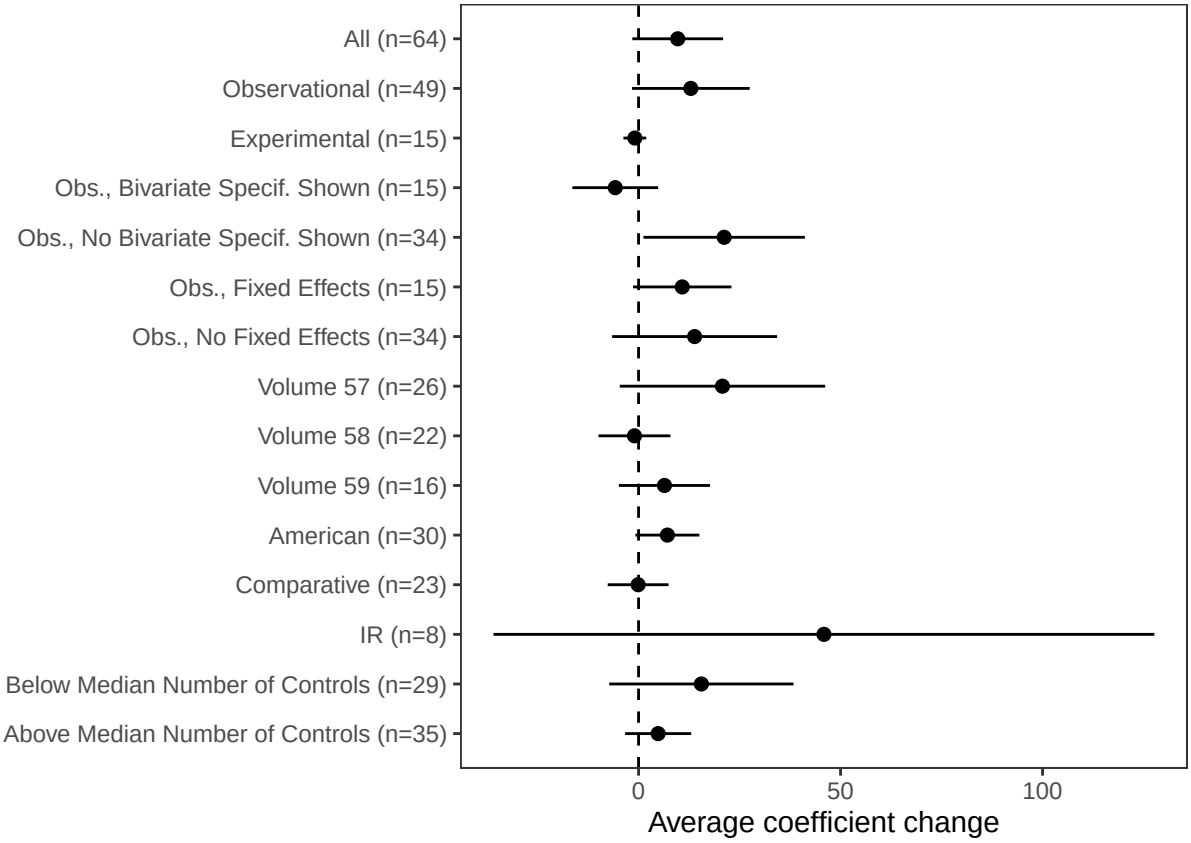


Figure S5: Changes in Coefficients that Do Not Show Bivariate

This figure is analogous to S2 but for coefficient changes instead of p-value changes. We calculate coefficient percentchanges with log plus one.

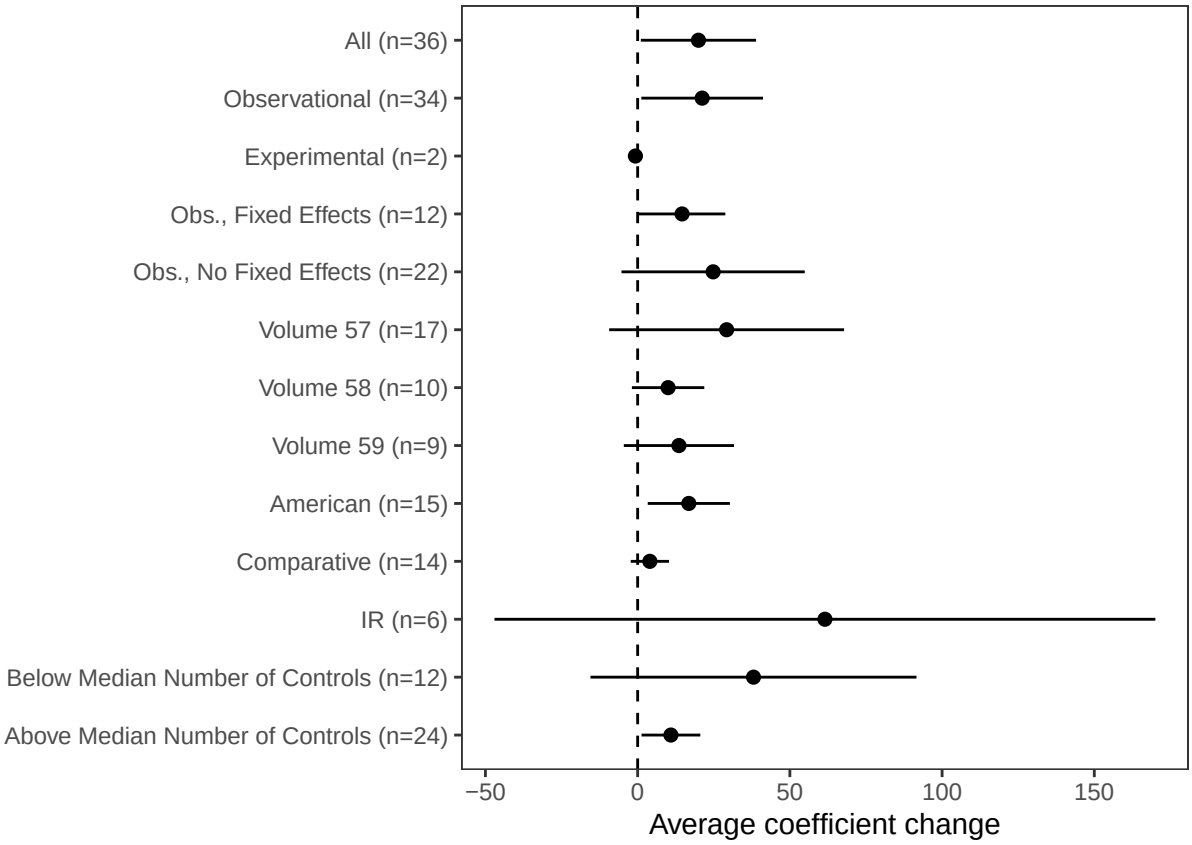


Figure S6: T-Value Changes in Observational Studies

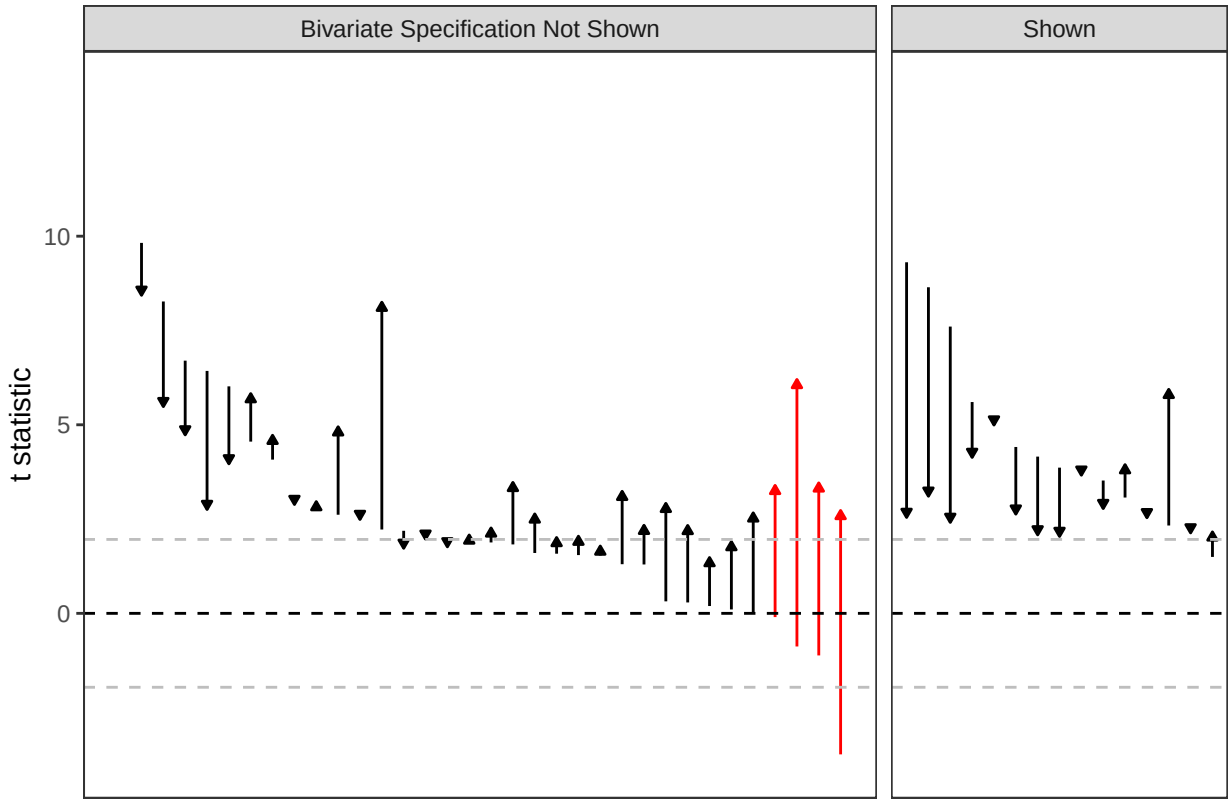


Figure S7a: Confidence Intervals of Bivariate and Full Specifications

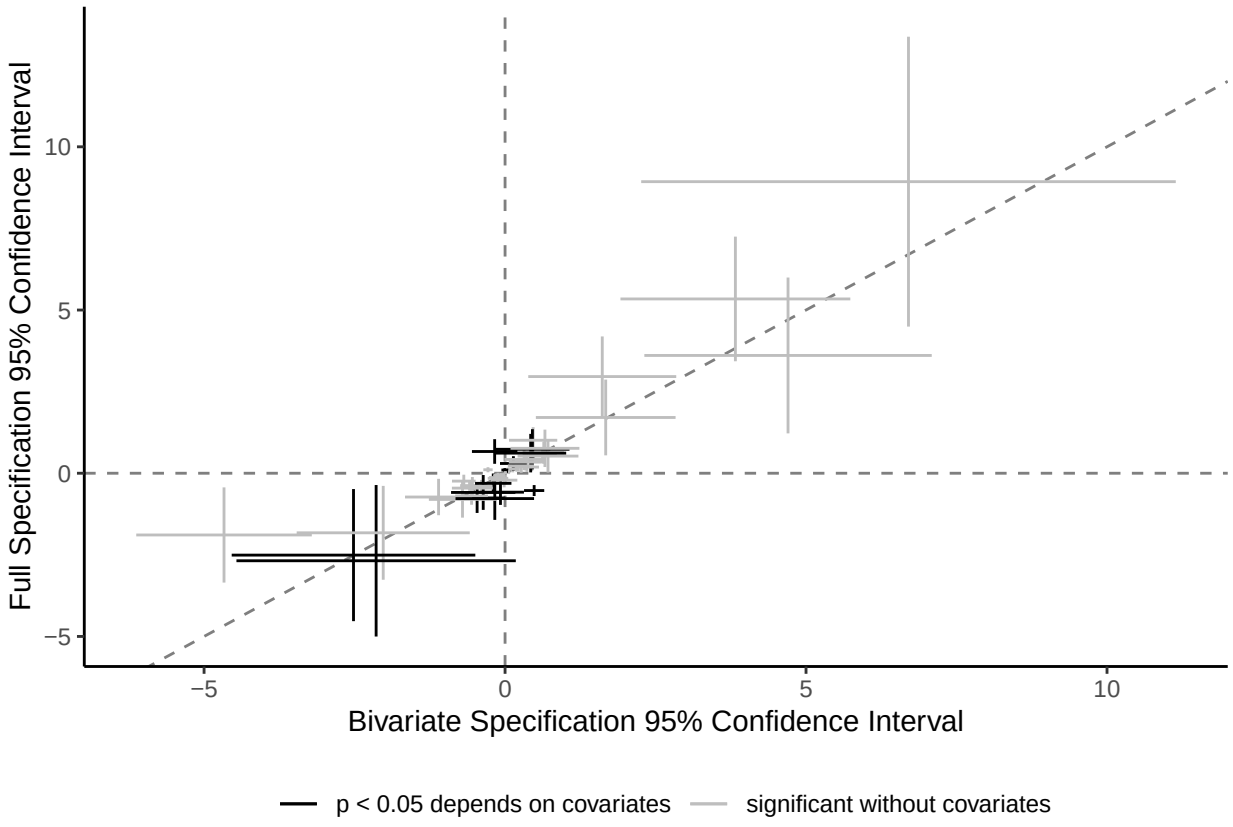


Figure S7b: Confidence Intervals of Bivariate and Full Specifications (Normed Absolute Value Coefficients)

Negative vales are from cases that switch signs. All variables recoded to 0-1.

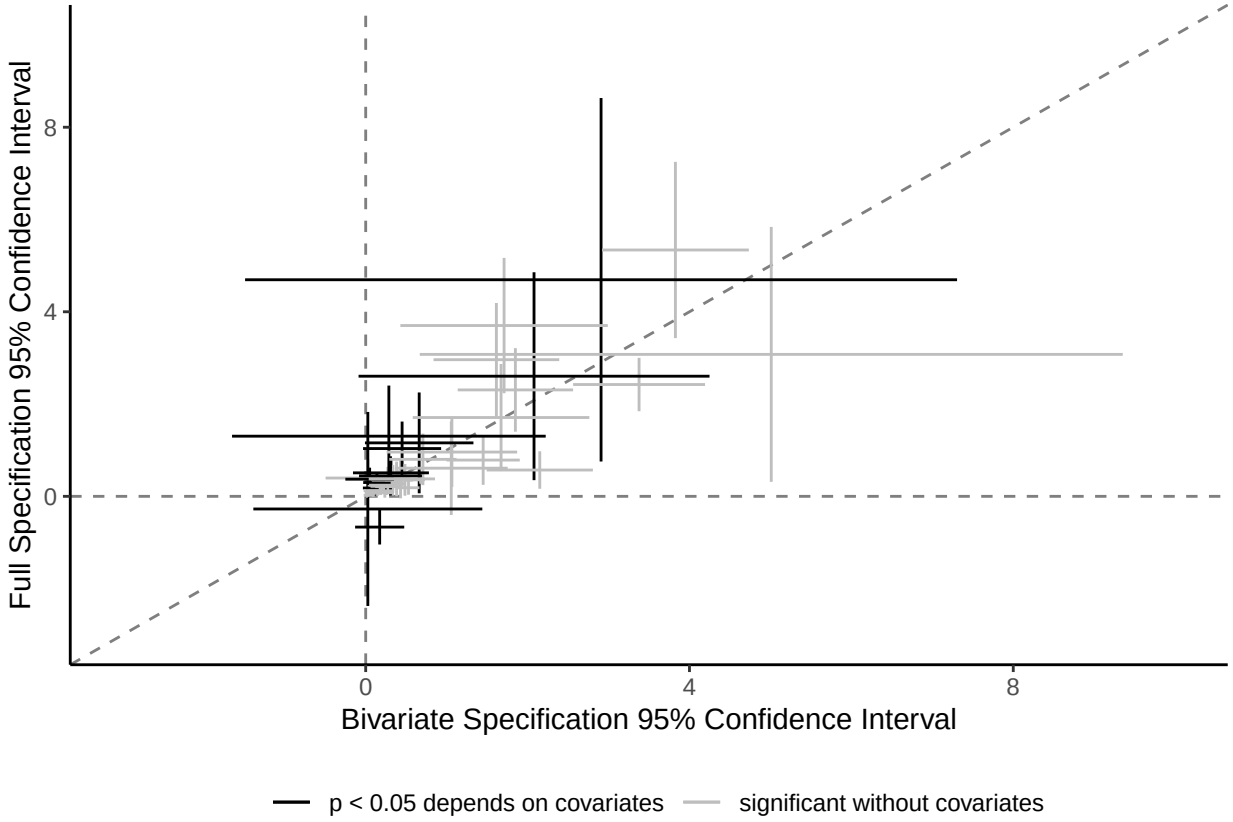


Figure S7c: Confidence Intervals of Bivariate and Full Specifications (Absolute Value Beta Coefficients)

All variables standardized so that their variances are 1. Negative vales are from cases that switch signs.

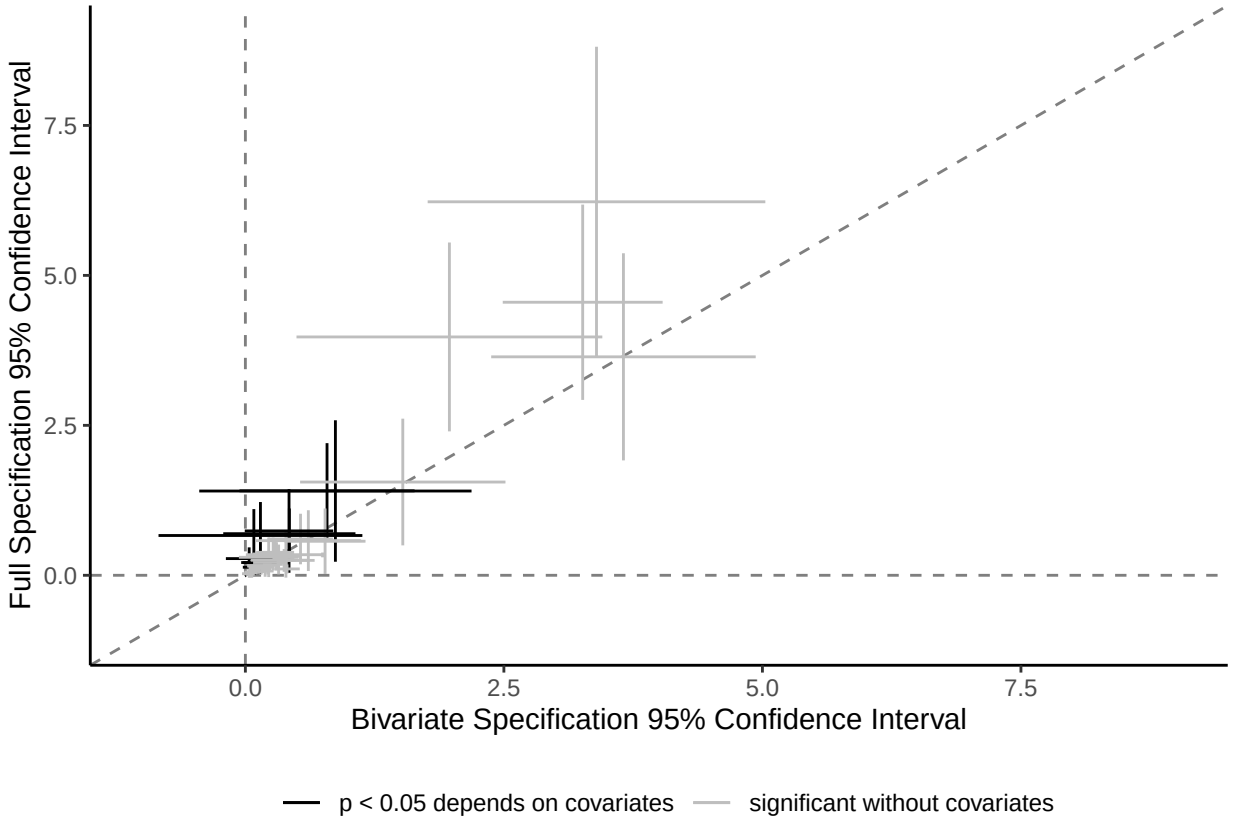


Figure S7d: Confidence Intervals of Bivariate and Full Specifications (Absolute Value Beta Coefficients–Zoomed)

All variables standardized so that their variances are 1. Negative vales are from cases that switch signs.

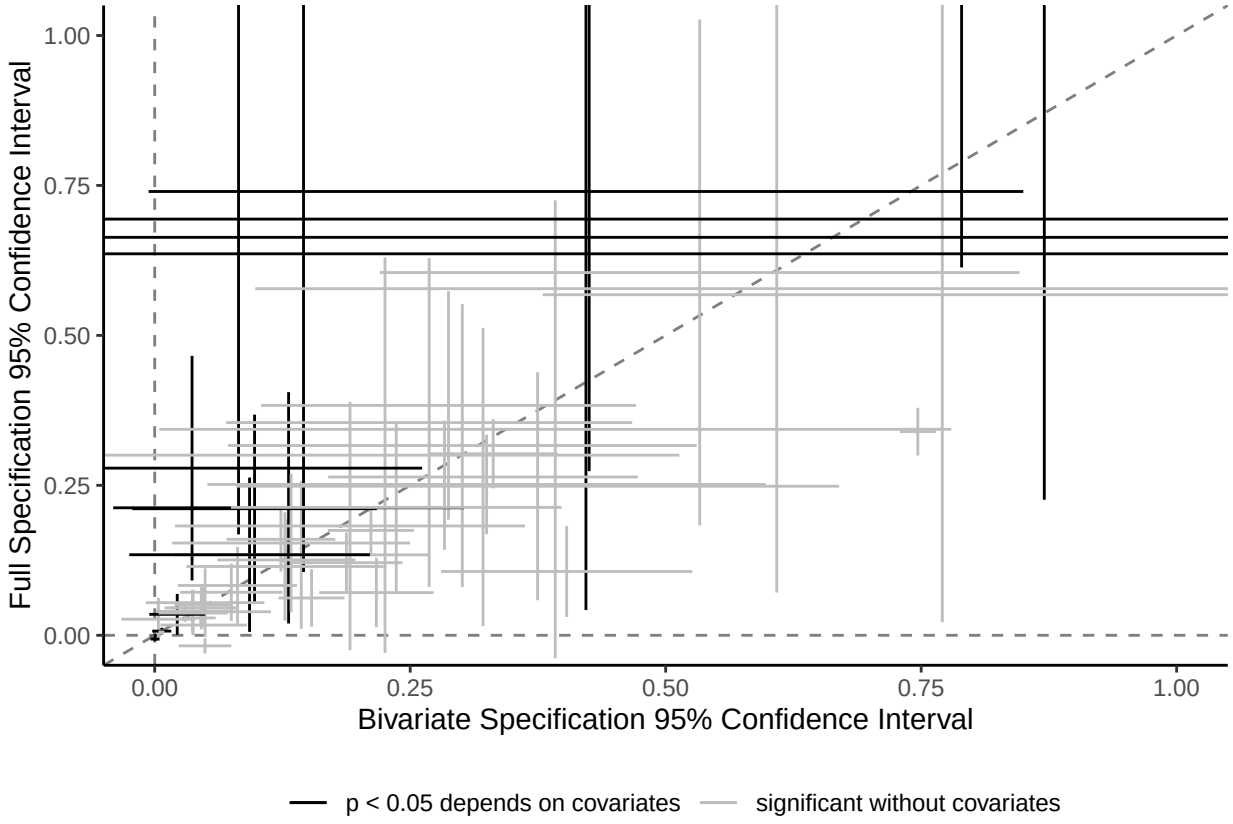
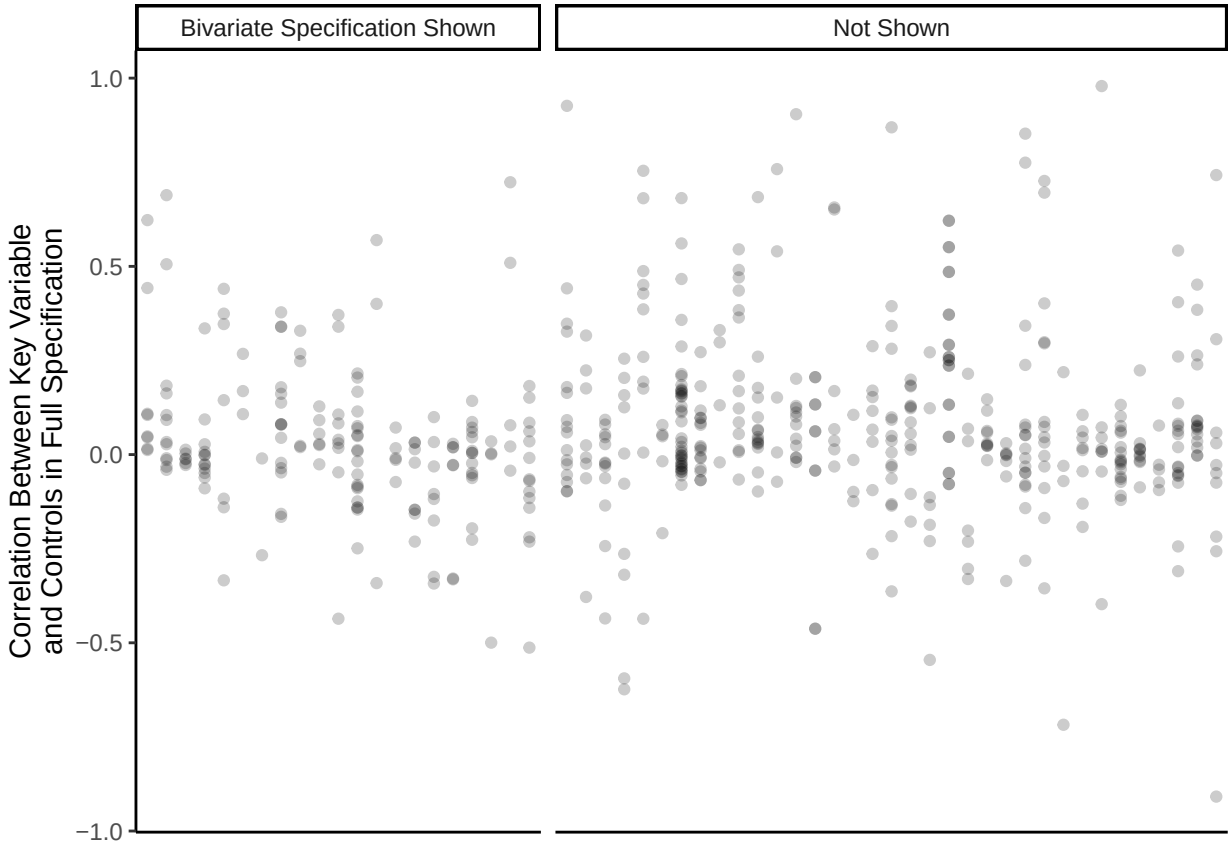


Figure S8: Multicollinearity in Observational Studies

The figure below shows the correlation between the key variable of interest and each control variable for every observational study. This figure follows the same design as Figures 2 and 3, which present our main findings, including the exact same ordering of studies. As we report in the paper, we find suppression effects helping authors achieve statistical significance on the far right side of this figure. If collinearity was contributing, we would expect to see higher correlations on the far right of this figure. As the figure shows, however, we see no such pattern.

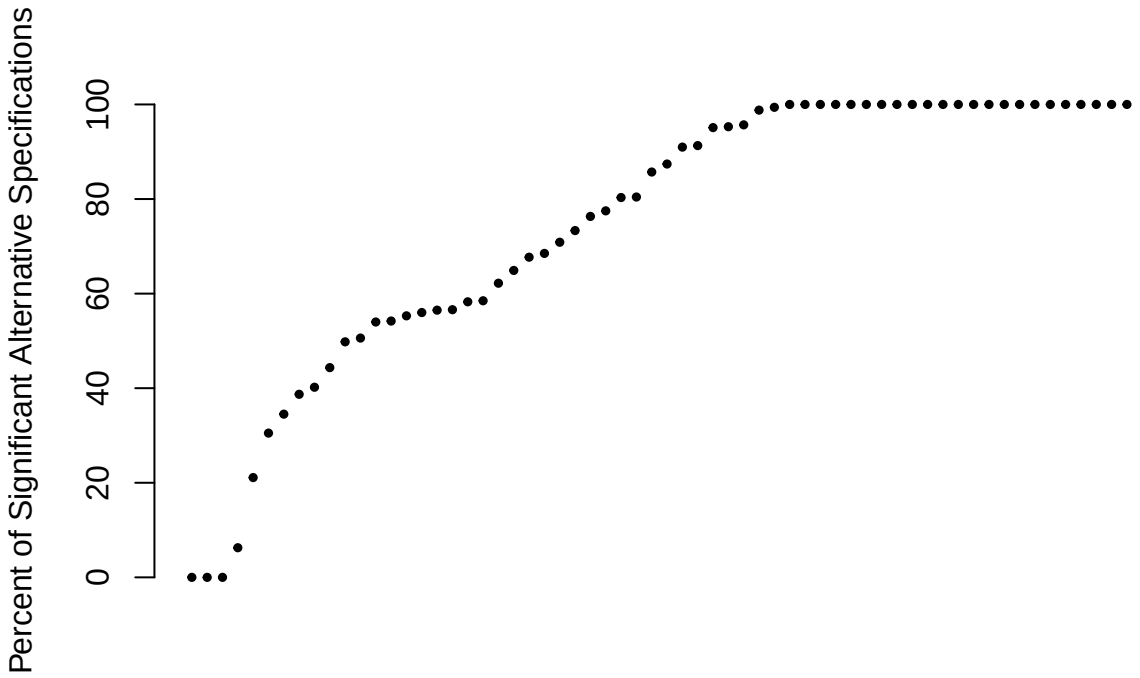


Specification Uncertainty

We conducted analyses to assess specification uncertainty from control choice, but concluded that these analyses were not especially informative. Here we present two of these analyses.

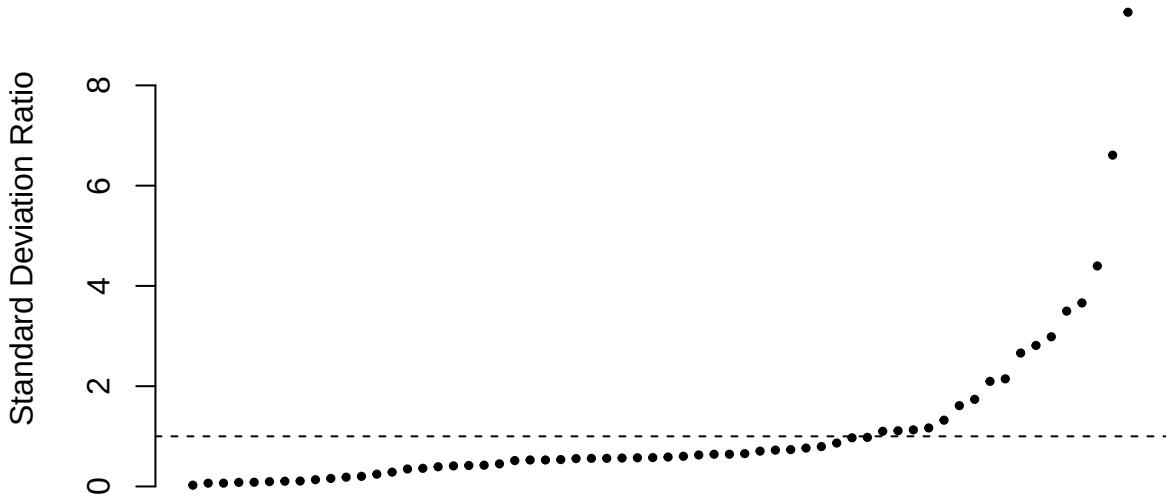
Statistically Significant in All Specifications

In the figure below, we present for each article the percent of all possible control combinations in which the key effect estimate was statistically significant (at the 0.05 level). Each dot represents one study and the studies are sorted by percent significant. The figure reveals that about a third of the studies were significant across all specifications. The other two thirds vary considerably with some being rarely statistically significant or never statistically significant. In studies with a large number of controls, we randomly selected control combinations.



Standard Deviation of Estimates

In the figure below, we present the standard deviation of the key variable's coefficient estimate across all possible control combinations. To scale them, we present a ratio: this standard deviation over the standard error of the key estimate (from the full specification). This ratio captures the degree to which specification uncertainty (from control choice) is larger or smaller than uncertainty reported by the articles' standard errors on the key estimates. Ratios of one, which are noted by the dotted line in the figure, show articles where the standard deviation across all possible estimates is about the same size as the standard error. If one believes that this standard deviation captures specification uncertainty, then ratios of one also imply that findings have about twice the uncertainty they report in their standard errors (see the previous page for why you should probably not draw this conclusion). The figure shows several articles that have truly massive amounts of uncertainty relative to the standard errors they report (as much as ninefold), but more than half have substantially less uncertainty from control choice than reflected in their standard errors.



Analyses Specified in Pre-Analysis Plan

Inflating with Controls

Plots of Key Coefficient under Full and Bivariate Specifications

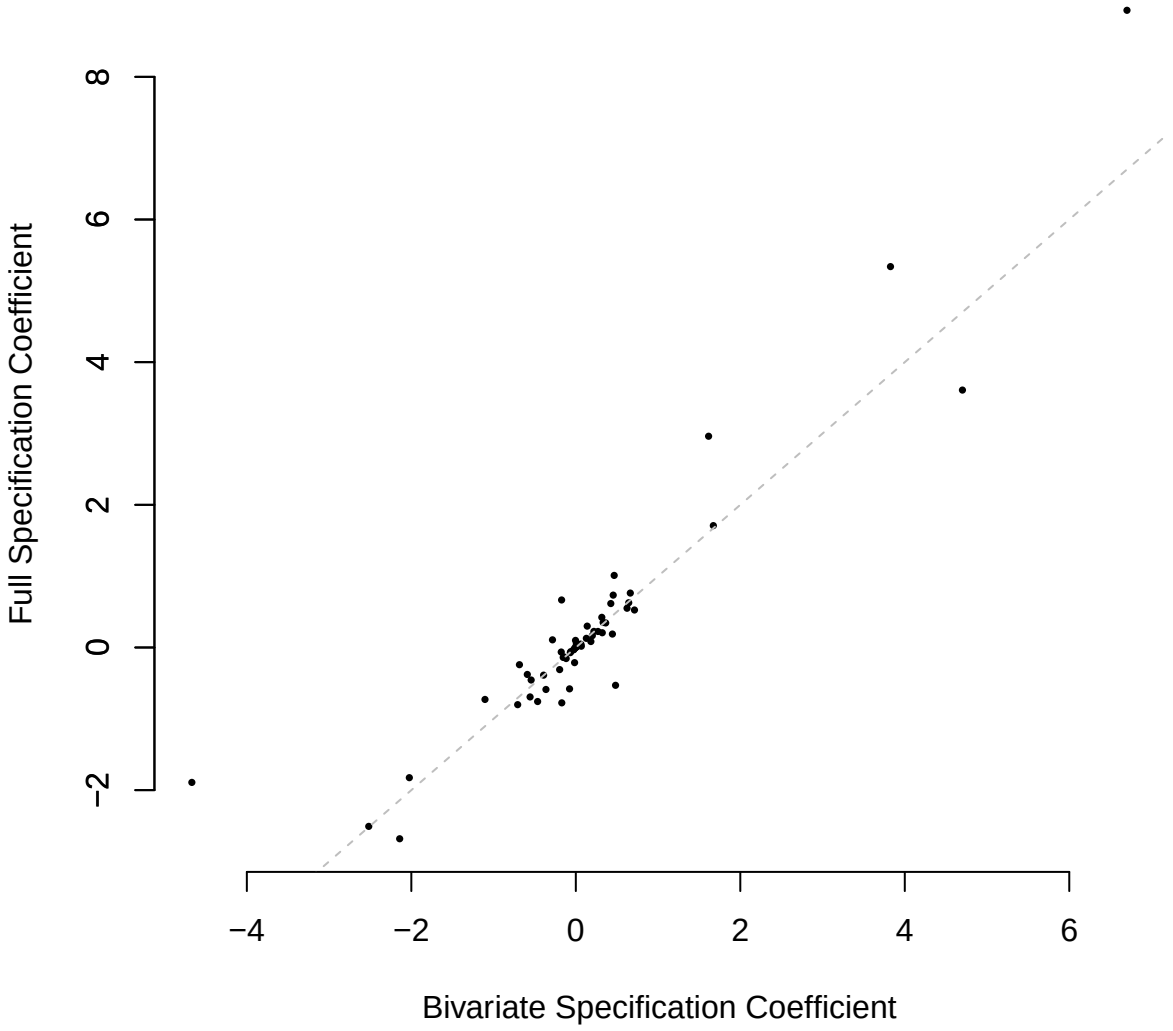
The first of set of statistics we will estimate is straightforward. How much does the inclusion of the controls used in researchers' main model change the main finding, compared to a model without controls? We will examine both the effect size estimate and the p-value. Often, researchers never report the estimate without controls (and we will code how often they do). This comparison will give us an overall sense of how often controls inflate estimates.

Since we did not know what the variation of the coefficients would look like, we prespecified three plots comparing the bivariate (often bivariate) and multivariate coefficients for each paper

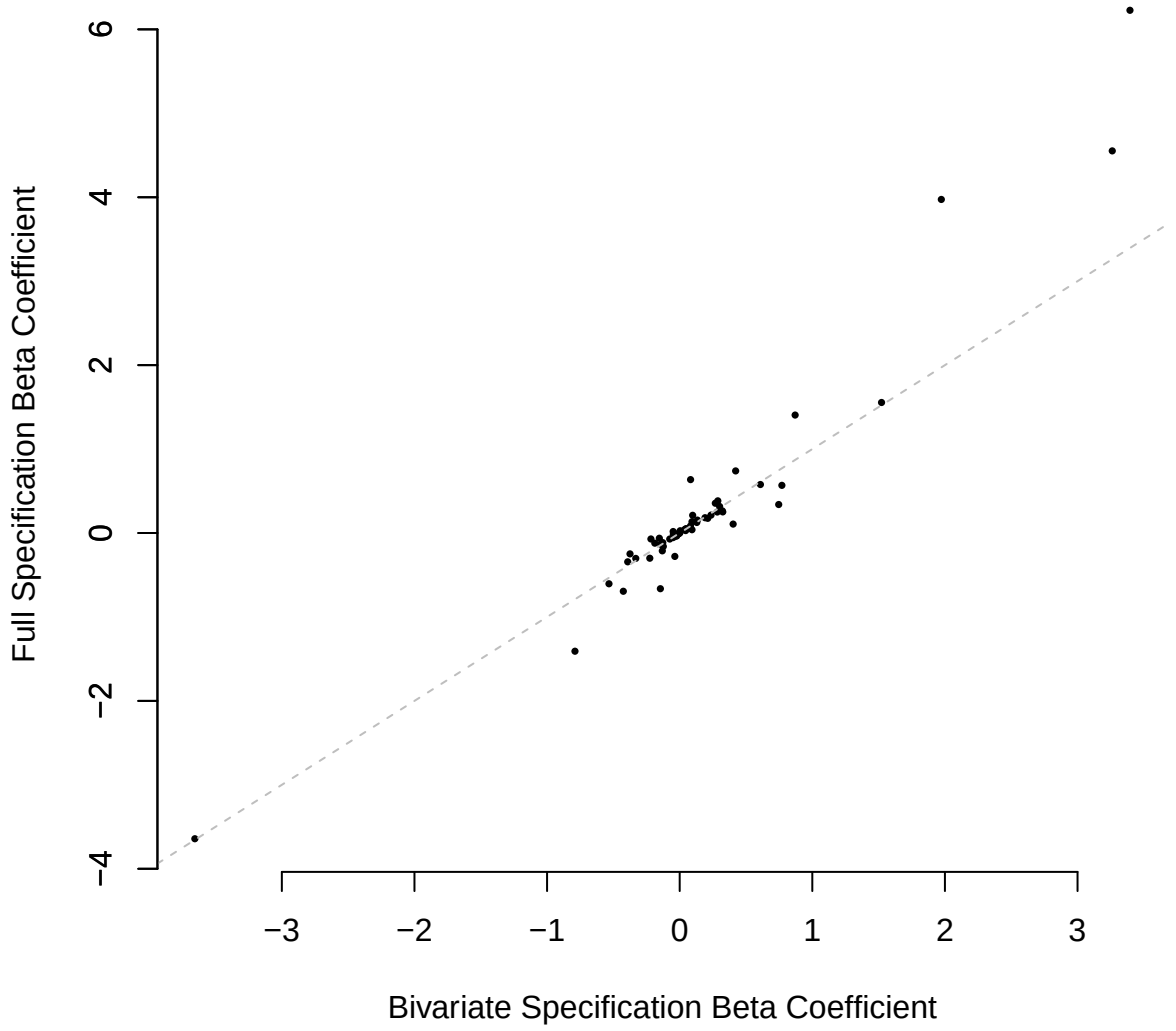
- The raw coefficients
- standardized beta coefficients, and
- the coefficients after rescaling variables to vary between zero and one.

The next three plots show those results. Unfortunately, the huge range and scales makes them almost impossible to interpret.

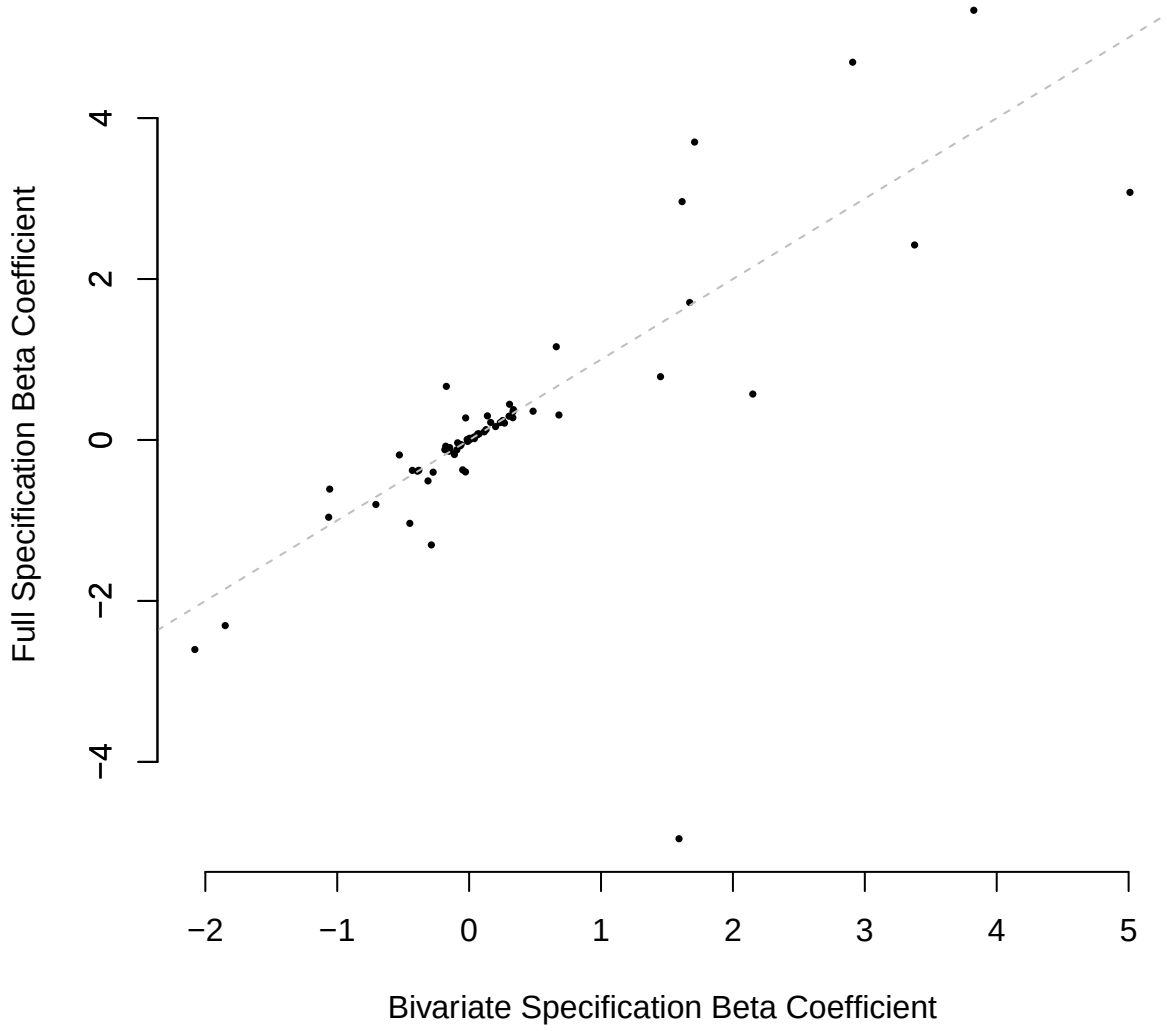
Key Estimate Under Full and Bivariate Specifications



Key Estimate Under Full and Bivariate Specifications (Beta Coefficients)



Key Estimate Under Full and Bivariate Specifications (Normed Coefficients)



Descriptive Statistics on Inflation

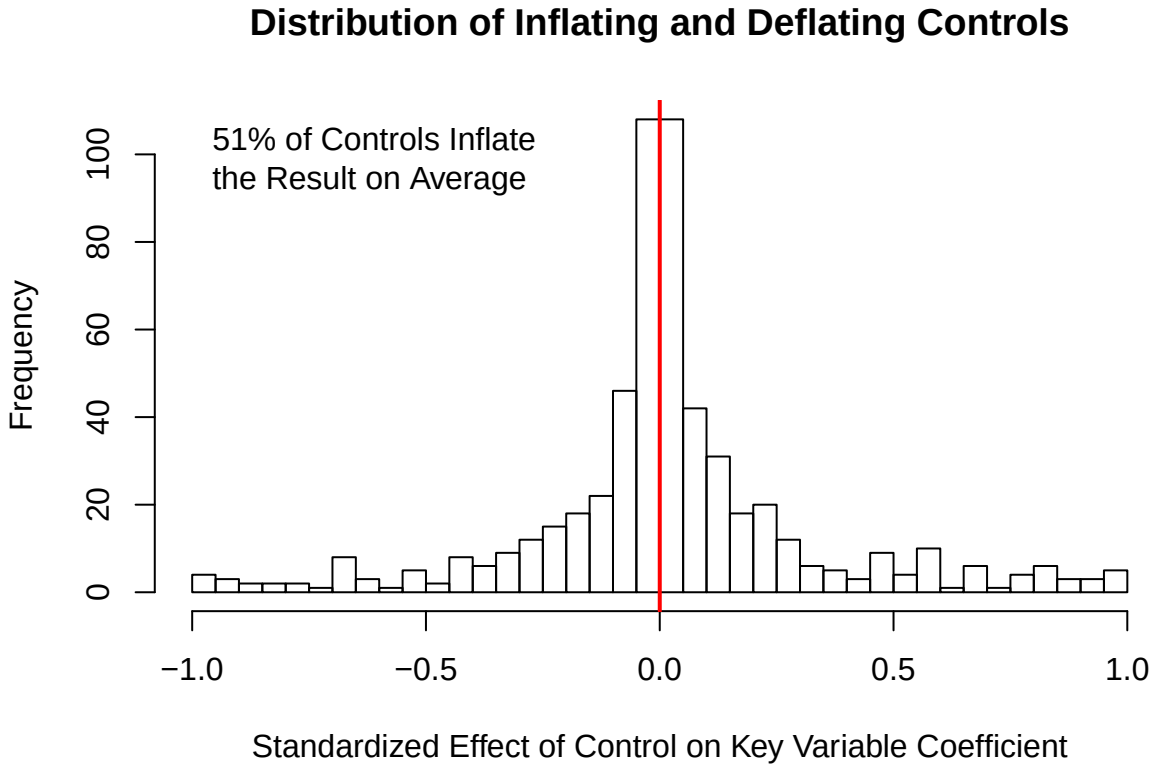
- Percent of studies where the coefficient of the key estimate under full specification inflates over the bivariate specification: 53% (34 of 64). Note that this includes all studies. When limited to studies where a bivariate specification is not shown, the percentage is 64%. When further limited to observational studies that do not show a bivariate specification, the percentage is 68%. When further limited to studies whose bivariate specification had a p-value greater than 0.05, the percentage is 100%.
- Percent of studies where p-value of key estimate does not achieve statistical significance under bivariate specification, but does in full specification: 25% (16 of 64). 69% of bivariate specification key estimates achieve statistical significance, 91% of full specification key estimates achieve statistical significance at $p<0.05$).
- Percent of models where the sign on the key estimate is different in the full and bivariate specifications 6% (2% where both estimates are statistically significant).

Sensitivity to Control Choice

- 98% of alternative specifications have a key effect coefficient of the same sign as the full specification.
- 74% of alternative specifications have a key effect coefficient of the same sign as the full specification that is statistically significant.
- 56% of alternative specifications have a smaller magnitude key effect coefficient size than the full specification.
- 33% of alternative specifications have a smaller magnitude key effect coefficient size than the full specification that is statistically significant.
- 46% of alternative specifications have a key effect coefficient size between full and bivariate specifications.
- 84% of studies have a minimum and maximum possible key effect of the same sign.
- The average standard deviation of estimates around the full specification across studies is 0.32.

Deflators and Inflators

- Percent of controls that inflate the estimate of the key variable on average; unweighted: 51%, weighted by study: 49%.
- Percent of controls that statistically significantly inflate the estimate of the key variable on average; unweighted: 39%.



Articles' Text about Controls

This section presents the text about controls from observational articles that did not present a bivariate specification.

1

Controls: Economic growth: Increasing income per capita is an important variable in our theoretical model as it increases support for welfare state spending, and is correlated with wage inequality. The percentage elderly can increase the generosity of welfare policy platforms, in particular on pensions, as the proportion of elderly in the electorate grows (Persson and Tabellini 2000). Demographic changes can also have an impact on wage inequality (see e.g. Hagemann and Nicoletti 1989 for a discussion of the effects of population aging on the labor market). Trade openness: implies a higher risk of income loss, and thus higher support for social insurance. Trade openness might also increase wage inequality (e.g. Wood 1994). Union density: There is a large literature on the impact of the unionization of the working class on the generosity of the welfare state (e.g. Korpi 1983). Encompassing unions also tend to equalize wages (Wallerstein 1999). The definition and sources of the control variables are described in Table A1. All variables are lagged one year, i.e. they refer to the situation the year preceding the election. Columns 1 and 3 present "stripped-down" models including the country fixed effects, the controls for the time trend, and the source dummies only. Columns 2 and 4 include con- trol variables. Control variables, and why we account for them in the regressions, are described in Appendix B in

the supporting information.

2

As controls, I use the same variables included in the matching stage. As a measure of judicial independence, I adopt the data provided by theCingranelli-RichardsHumanRights Data Project (CIRI; 2009; JUDICIALINDEPENDENCE), which are coded 0 for not independent, 1 for partially independent, and 2 for generally independent. I include ameasure of regime type using the Polity IV data (Marshall and Jaggers 2002; POLITY) because democracies aremore likely to respect human rights (Davenport 1995, 1999; Poe and Tate 1994; Poe, Tate, and Keith 1999). Newer regimes and well-established regimes may have different preferences, so I control for this factor using the Polity IV data (Regime Durability). Foreign wars and civil wars may result in periods of increased repression (Fariss and Schnakenberg 2014; Hill and Jones 2014; Schnakenberg and Fariss 2014). Civil wars, in particular, may result in periods of lawlessness during which even powerful legislatures have a diminished capacity to con-strain the other branches of governments. I use data from the UCDP/PRIO armed conflict database. NGOs play a key role in political mobilization against oppression and may succeed in improving government practices. I include the number of international NGOs (INGOS)in a country using the data provided by Hafner-Burton and Tsutsui (2005). Economic development is a well-known predictor of human rights practices (Henderson 1991; Poe and Tate 1994; Poe, Tate, and Keith 1999), and I control for this using a measure of per capita gross domestic product (GDP) provided by the World Bank. I use the natural log of this measure because this effect is likely nonlinear (Davenport 2007a). To address potential differences among states of different sizes and potential monitoring biases based on this factor, I follow much of the literature in including the natural log of a state's population, using data provided by the World Bank.

3

For our first set of analyses, we control for several party and country attributes that might be argued to in- fluence the extent to which our hypothesized relation- ships hold. First, for individual parties, we control for Party age (measured in years), Party size (measured as a party's share of the national-level vote), and whether or not a party is a Niche party (i.e., ecologist, communist, or nationalist party). At the country level, we control for New democracy, the Effective number of electoral parties (ENEP), and Economic performance.

4

In addition to the above variables of interest, the models include five sets of variables that measure respon-dents' socioeconomic status, perceived commonalities with Latinos, group identification, religious indicators, political knowledge, basic demographics, and contex-tual indicators. 13 Extant research commonly notes Latino immigration attitudes are associated with these factors (Binder, Polinard, and Wrinkle 1997; Branton 2007; Carey, Branton, and Martinez-Ebers 2013; Citrin et al. social and economic integration (as measured by commuting to work) with the urban core (U.S. Census Bureau 2012, 5). 12 This measure includes only the respondents who lived in an area where at least one protest occurred before the respondent was surveyed. 1997; Hood et al. 1997; Rocha et al. 2011; Sanchez 2006). These variables are important in controlling the demo-graphic and political

differences of respondents between the treatment and control groups. The assignment of the respondents to the treatment and control groups was al-most random, yet the respondents' decision toparticipate in the survey might be influenced by their personal char-acteristics and exposure to the rallies. Indeed, we find significant differences in the demographic and political features of the respondents in the treatment and control groups. Accordingly, these control variables play an im-portant role in isolating the effects of the rallies on the outcome variable. Respondents' socioeconomic status is measured by their educational background, income level, andpersonal financial situation. Education is measured by an 8-point scale ranging from 0 to 7. Respondents' income is mea-sured by four variables that equal 1 if respondents belong to one of the quartiles on the scale of income and 0 other-wise. The lowest quartile serves as the baseline category. Additionally, we include an indicator variable that equals 1 if respondents refused to report their income status and 0 otherwise. Respondents' assessment of their personal financial situation is measured by a 3-point scale rang-ing from (1) gotten worse, (2) stays about the same, and (3) gotten better. Respondents' perceived commonalities with other Latinos is measured by an item asking, Thinking about issues like job opportunities, education or income, how much do [ethnic subgroup] have in common with other Latinos or Hispanics? Would you say [ethnic subgroup] share a lot in common, some things in common, little in common, or nothing in common with other Latinos? The measure is coded (1) nothing, (2) little, (3) some, and (4) lot. Tomeasure group identity, we use a survey item that asks respon-dents to choose the category that best de-scribes them: Latino/Hispanic, American, their national-origin group, andnone of these.We created twomeasure to capture their primary group identity. For the firstmea-sure, a value of 1 was assigned to respondents who chose their national-origin group and 0 otherwise. The second measure assigns a valueof 1 to respondentswho identified more strongly with their American identity and 0 oth-erwise. The baseline category reflects a Latino/Hispanic primary identification. To measure a respondent's political knowledge, we rely on an additive measure based on responses to three survey items.Respondentswereaskedto identify theparty that had a majority in the House of Representatives, the presidential candidate whowon the 2004 election in their state, and the political party that is more conservative than the other. The responses were coded 1 if a response was correct and 0 otherwise. Scores range from 0 to 3, with higher scores indicating a higher degree of political knowledge. To account for the impact of religion and civic participation, the models include three measures: Catholic,Attend Church,andCivic Participation.Catholic is a dichotomous measure that indicates whether a respondent is Catholic.Attend Churchrepresents regular church attendance, which is a dichotomous variable coded 1 if a respondent indicates she attends religious services once a week or more.Civic Participationis a di-chotomous measure that indicates whether a respondent is involved in a civic organization. Given the importance of the Catholic Church and civic organizations in activist efforts—not only with regard to the 2006 rallies but also more generally (Schmidt et al. 2008)—we include this to account for the potential influence of involvement in the church and civic organizations on immigration attitudes. We control for differences in national origin using dummy variables for country or region of origin/descent. Thedummyvariables equal 1 if respondentsdescend from Mexico, Cuba, Puerto Rico, the Dominican Republic, Central America, and South America. Mexican respon-dents serve as the baseline category. Further, we control fordifferences inage-

measuredbyyears andgender,which equals 1 if the respondent is female and 0 otherwise. Finally, our model includes aggregate-level measures of the socioeconomic and ethnic context. Socioeconomic context ismeasured by tract-level percent high school ed-ucated. Ethnic context is measured by the percentage of Latinos at the tract level. These measures are included to account for variance in immigration attitudes attributable to the context in which one resides.

5

Seven control variables capture country- and election-specific factors that may influence the probability of a democratic transition. First, it is likely that democratization occurs primarily in elections that deter-mine who will hold executive power. Because the stakes are higher in these contests, they are more likely to elicit higher levels of domestic mobilization for democracy. I include a dichotomous variable, main election,coded as a 1 for presidential elections in presidential (or mixed) systems and legislative elections in parliamentary systems (Simpser and Donno 2012). Second, I include a variable indicating whether theincumbent was running in the election, which is expected to decrease the chances of de-mocratization (Cheeseman 2010; Maltz 2007).15 It is also important to account for the country's previous experi-ence with elections. Lindberg (2006, 2009) argues that holding elections fosters institutional change and greater respect for civil liberties and that these changes cumulate over time. Relatedly, Bratton and Van deWalle (1997) ar-gue that the more elections held under authoritarianism, the greater the chances for democratization. I thus in-clude a variable that sums thenumber of previous elections held under a continuous authoritarian spell. 16 Two variables control for economic conditions: first avariableforGDP per capita,lagged one year. 17 High income is reliably associated with democracy, though its effect on democratic transitions is less clear. Przeworski et al. (2000, chap. 2), for example, find that dictatorships at high (but not the highest) levels of income are more likely to democratize. Second, I include a variable for GDP growth, measured as the percent change in a coun-try's GDP from yeart-2tot-1.Ifitistruethatgood eco-nomic performance bolsters authoritarian regimes, the coefficientonthisvariableshouldbenegative

6

Our analyses included a standard set of demographic and political control variables, as well as additional theo- retically relevant contextual and individual-level controls. First, we control at the county level for the Median House- hold Income, Percent Black, Total Population, and Bush Vote 2004 to account for the distinct effects of variance in absolute economic conditions, racial composition, county size, and political culture across respondents' county of residence.7 At the individual level, we include standard controls for Education, Age, Gender (1 = male), Employ- ment Status (1 = unemployed), Union Membership (1 = respondent and/or spouse is union member), Party Iden- tification (5-point scale; 5 = "strong Republican"), Ide- ology (5-point scale; 5 = "very conservative"), and Re- ligious Attendance. We control for unemployment and union membership as standard controls in research an- alyzing attitudes toward the economy and government redistribution (e.g., Anderson and Pontusson 2007; Fong 2001; Scheve and Slaughter 2004), and for religious at- tendance because religiosity has been linked to general ideological orientations, such as humanitarianism and egalitarianism (Bartels 2008; Feldman and

Steenbergen 2001) and social welfare policies (Sears et al. 1997).

7

Alternatively, it is possible that towns with casualties tended to be environments of lower socioeconomic sta-tus. Though the differentiation in induction rates across income classes is disputed (Flynn 1993; Zeitlin, Lutter-man, and Russell 1973), poor and working-class men are significantly overrepresented in combat-related deaths (Kriner and Shen 2010; Zeitlin, Lutterman, and Russell 1973). Table 4 reports estimates from probit analyses that include socioeconomic and political contextual variables as well as interaction terms between each contextual vari-able and the binary indicator for a low lottery number. The first two columns include contextual-level socioeco-nomic indicators. Binary variables indicate whether the town population is above the median for the sample, the town's median household income is above the sam-ple median, and the percent of residents who did not graduate from high school is below the sample median. Columns 3 and 4 report results of probitmodels inwhich binary indicators for above-median turnout in 1968 and above-median shares of Democratic vote share in 1972 are added. Results indicate that accounting for local po-litical or socioeconomic attributes does not substantively change the estimates of the interaction effect of a son's as-signment to high draft priority status within the context of a prior local war casualty.

8

Throughout the analysis, a number of other variables are used as controls. All are described further in the sup- porting information (item 1). Intensity of the trafficking problem in countries of origin, transit countries, and desti- nation countries was generated based on the 2006 United Nations Trafficking in Persons report and is a constant for all years. Missing information measures the availabil- ity of information on trafficking, which may influence the ability of the United States to include a country in the report in the first place. It is a count of how often a country has missing information on 10 types of data in a given year, including seven unrelated to trafficking. To reflect the U.S. State Department's access to information about trafficking in the country, we also created a variable, NGO density, based on the number of NGO mentions in the U.S. TIP reports, extended backward to all years, creating a constant measure for almost all countries included in the analysis. Finally, regional density of crimi- nalization measures the proportion of countries within a country's region that had criminalized as of the previous year. Other variables include civil liberties from Freedom House, an indicator of 2000 TIP protocol ratification, to-tal population (logged), as well as measures of a country's bureaucratic quality, rule of law, corruption, or the share of women holding seats in parliament. All sources and mea- surement details are listed in the supporting information. .....[after analysis] Other factors also matter. Countries are more likely to criminalize the greater the share of women in parliament, the greater their civilliberties, the greater the regional density of criminalization, and if they have ratified the 2000 TIP protocol. Interestingly, U.S. aid in itself appears to have little effect on criminalization.

9

These effects are notably resilient to the effect of issue position extremity. Issue position

extremity does significantly increase thermometer bias, but farmoreweakly than sorting or partisan identity strength. Controlsare included for education, sex (dummy), white race (dummy), age, southern residence (dummy), urban residence (dummy), frequency of church atten-dance (as ameasure of religious commitment), and evan-gelicalism (as a measure of religious conservatism, a dummy variable based on denomination). All continu-ous variables are coded to range from 0 to 1

10

Allmodels control for basic demographics, including educa-tion,age,andgender.

11

In addition to these model-specific control variables, we include the number of standing committees to which a bill is referred, which we expect to have a positive impact on the level of legislative scrutiny, and thus the number of changes made to government bills. We also control for the (logged) size of the bill introduced, since bills with many articles are naturally going to have more articles changed, on average, than bills with few articles. We also include a measure of the number of days a bill spends in the leg- islative process, as well as an indicator for whether a bill expires before the plenary vote. We also include an indica- tor for the numerical status of the government to account for the possibility that minority governments may have to make more policy changes than majority governments to entice opposition parties to support legislation. Finally, we include indicators to account for country-specific and issue-specific fixed effects.

12

In Model 2, we control for evaluator (Ideol- ogy of Eval.) and target ideology (Ideology of Target; both continuous measures), sex, and average attraction levels.

13

A vector of variables, includes the mean, maximum, and minimum ability scores of all representaives, who form the entire pool of potential candidates, and Xjs is a set of group controls as in Equation (5). To account for the nested na-ture of the data, we include(2) , a random intercept for regions. we also include the manager's ability and the followinggroup-level controls (Xjs): number of group members, age of the group, and a measure of ethnic fractionaliza-tion.26To account for the nested nature of the data, we include b2 , a randomintercept for regions;finally, js isthe residual error term. table note : Controls, centered on their mean values, include the number of association members (in units of 50), the DC age, and its ethnic homogeneity (ELF) using a simple Herfindahl concentration index

14

The most serious potential confound for the analysis be-low is distinguishing relative political importance and relative population. Similarly sized ethnic groups are thought to be more likely to clash (Buhaug, Cederman, and Rød 2008; Horowitz 1985; Montalvo and Reynal-Querol 2005; Reynal-Querol 2002) and are also likely to be similar in political importance. I control for: Demographic polarization=n 2 i nj +nin 2 j , (1) where ni and nj

are the population shares of the plurality group in the enclave and the group opposed to state-hood, respectively. 27 Esteban, Mayoral, and Ray (2012) argue that cultural distance exacerbates the effects of po-larization. I calculate cultural polarization as suggested by Fearon (2003): Cultural polarization= n 2 i nj +nin 2 j dij, (2) where dij is the linguistic distance between the enclave plurality and the opposing language group, normalized to fall between 0 and 1. 28 Additional Confounds Other confounds are variables that may influence political importance to the Congress and violence. If grievances caused groups to both vote against the Congress and use violence, relative INC representation might be cor-related with violence by virtue of proxying for dissatisfaction. Therefore, I control for the absolute level of Congress representation of statehood proponents (Ln enclave plurality group's INC rep.). Other likely correlates of violence plausibly related to political impor-tance are population (Ln enclave plurality group's population); economic development, measured as the share of the workforce in agriculture (Agricultural labor sharin enclave); and distance to the capital (Ln km to New Delhi). 29 Regional inequalities were of limited salience dur-ing reorganization because states levy few taxes. How-ever, states do have authority to tax and redistribute agri-cultural holdings. Demand for land reform may be an important control, therefore. Landless rate in enclave is the share of the agricultural workforce that is landless. Finally, I also measure enclaves' Hindu population share (Hindu share in enclave). Wilkinson (2008) and Capoccia, Saez, and de Rooij (2012) suggest that religious disputes in India have been particularly violent. Brass (1974) argues that partition made New Delhi wary of territorial demands construed in religious terms. Since 29 Data on population, sector of employment, landholdings, and religion are from Census of India (1951). religion and voting patterns are also correlated, religion is a potential confound. ... I control for polarization, the enclave-plurality group's INC representation, and population, all variables that might be proxied by relative INC representation. Also, the literature suggests that New Delhi was least accommodating of religious minorities and movements posing a separatist threat; therefore, I control for religious composition and distance to New Delhi.

15

Of course, controlling for potential confounders is important. Personalism might, for example, covary with geographic factors in a way that obscures the relationship between regime type and pursuit of nuclear weapons. Similarly, personalismcouldbe associat-edwith economic development or military capabilities (in the sense of the Correlates of War material-resources index), something thatmay be associatedwith a greater likelihood of pursuit of nuclearweapons. Inviewof these concerns,we estimate logistic regression models that control for confounding variables, but keep in mind the importance of limiting the inclusion of "posttreatment" variables that would bias our estimates of the effect of personalist regime type. Since observations over time within a particular country are clear-lynot independent, failure toaccount for temporal dependencewithineachcross-sectioncan result in underestimates of standard errors, leading to unduly optimistic inferences (Beck, Katz, and Tucker 1998). We thus include three regressors to model time passed with-out the pursuit of nuclear weapons:t, t 2 ,andt 3 (Carter and Signorino 2010).

16

Verified voter registration status and a full set of demographic controls are available for 2,249 of the CCES respondents, but the results are nearly identical when the analysis is expanded to include all 4,435 respondents who reside in one of these districts.

17

Demographic control variables. We include age be- cause older respondents may be more likely to hold more negative views of government. Education is controlled for because the literature suggests that as levels of educa- tion increase, respondents may possess a higher sense of efficacy and trust in government. We also include gender to examine differences between Latinos and Latinas. The second set of variables is related to generational status, time in the United States, and national origin groups. The models include a dummy variable for whether respondents are first generation, since they are less likely to be politically acculturated and may be more likely to find politics confusing. Given differences in Latino political attitudes based on national origin groups (Abrajano and Alvarez 2010; Alvarez and Garcia-Bedolla 2003), we also include dichotomous measures for whether respondents are of Mexican, Puerto Rican, Cuban, Do- minican, or El Salvadoran decent.12 For example, Cubans are usually the most conservative among national origin groups, whereas Mexicans and Puerto Ricans are typically the most liberal. We also include a dummy variable for respondents who completed the instrument in Spanish, Spanish Pref, in the event that Spanish-dominant speakers vary in their attitudes. Similarly, we utilize a variable that measures the percent of a respondent's life spent in the United States.13 Finally, we also include three variables related to me- dia consumption because the source of news on poli- tics could shape how respondents viewed the marches and their relationship with government. We include a measure for how frequently respondents watch the news on television. Second, we utilize a question asking how often respondents read a daily newspaper. We created a dichotomous variable indicating whether people rely

18

The frequency with which a representative speaks on the floor against the war in Iraq. Most importantly, Republi- cans are significantly less likely to criticize the war policies of a Republican president than are Democrats. Moreover, even after Obama's ascent to the White House, after five years of public support for the war, Republican mem- bers should be significantly less likely to criticize it than their counterparts across the aisle. As a result, the model first includes a dummy variable identifying Republican members of the House. Moreover, to assess whether the effects of constituency casualties are different for Democratic and Republican members, the model also includes the interaction of this Republican dummy variable with the local casualties measure described previously. A member's position within the chamber hierarchy might also affect her or his willingness to use the insti- tutional forum of a floor speech to criticize the war. To control for the possibility that members of the leadership might be more or less willing to attack the White House's policies, we include a dummy variable identifying those holding leadership positions. Alternatively, members of committees dealing with foreign affairs and intelligence might be more willing to defend Congress's institutional prerogatives and confront the executive's handling of mil- itary affairs. To account for this possibility, we include a count of each

representative's memberships on the for- eign relations, armed services, intelligence, or homeland security committees. As a final control for institutional context, we include a measure of each member's seniority within his or her chamber. More senior members who are more invested in their institution may be more willing to confront the executive branch in the military arena; moreover, particularly in the House where floor time is more tightly regulated, more senior members may simply have more opportunities to express their opinions on the war than their junior colleagues (e.g., Hall 1996). A significant literature on civil-military relations sug- gests that veterans of the armed forces may view military matters differently than civilians (Dempsey 2010; Feaver and Kohn 2001; Gelpi and Feaver 2002). Accordingly, we include an additional variable, whether each House mem- ber had served in the armed forces, to the models. Given the importance of members' personal backgrounds in influencing their voting behavior (Burden 2007), we also included a series of demographic variables to identify each member's race and gender (Dodson 2006; Rocca, Sanchez, and Nikora 2010). Finally, because this first model pools data from multiple Congresses, it also includes dummy variables for the 109th, 110th, and 111th Congresses.

19

Political Predispositions. We control for several factors. Two questions commonly included in the American Na-tional Election Studies (ANES) survey tapping egalitarianism (i.e., worry less about equality and gone too far with equal rights) were combined to form a weak scale (= .48, r = .31). Two standard ANES items tapping individualism (i.e., blame self if don't get ahead and poor because they don't work hard) were also combined to form a scale (= .49, r = .32). Both individualism and egalitarianism were rescaled to range from 0 to 1 (individualism: M = 0.41, SD = 0.27; egalitarianism: M = 0.55, SD = 0.29). Political ideology and party identification were measured using the standard ANES 7-point format, recoded to vary from 0 to 1. We estimated all models with a variable corresponding to the stage of interview, which does not change the substantive results. The cooperation rate was calculated as the ratio of completes to completes, partials, and refusals. The overall response rate was 31% calculated as the ratio of completes to completes, partials, refusals, and no answers for numbers that were clearly households. A maximum of 15 attempts were made at each number. Interestingly, the two scales were virtually uncorrelated (r = -.02). This can be attributed to the fact that the items were worded in different directions; previous research also finds low zero-order correlations between positive and negative stereotype items in the absence of statistical corrections for systematic measurement error (which cannot be performed here due to the small number of items available in our survey; see Levine, Carmines, and Sniderman 1999). However, the correlation is more in line with expectations among low self-monitors (i.e., among respondents scoring below the 25th percentile of the self-monitoring scale), at r = .13 (for high self-monitors, i.e., those scoring above the 75th percentile of the scale, the correlation is negative, at r =-.18). 1, with higher scores indicating greater conservatism and identification with the Republican Party (ideology: M = 0.49, SD = 0.32; party identification: M = 0.48, SD = 0.34). We also control for age, measured in years,4 education (1 = bachelor's degree or greater, 0 = otherwise), and gender (1 = Female, 0 = Male).

20

All of the models include controls for the incumbent party's vote share in 2001 (Incumbent Vote 2001), whether the incumbent candidate was running (Incumbent Candidate), which party was the local incumbent party (MMD Incumbent,UPND Incumbent), the number of candidates for the parliamentary seat (Number of Candidates), and the incumbent candidate's years of experience (Years Since Incumbent First Elected).

21

Why do some individuals engage in more religious activity than others? And how does this religious activity influence their economic attitudes? We present a formal model in which individuals derive utility from both secular and religious sources. Our model, which incorporates both demand-side and supply-side explanations of religion, is unusual in that it endogenizes both an individual's religious participation and her preferences over economic policy. Using data on over 70 countries from the pooled World Values Survey, we find that religious participation declines with societal development, an individual's ability to produce secular goods, and state regulations on religion, but that it increases with inequality. We also find that religious participation increases economic conservatism among the poor but decreases it among the rich. Our analysis has important insights for the debate about secularization theory and challenges conventional wisdom regarding the relationship between religious participation and economic conservatism.

22

The likelihood that Congress accommodates the president's requests depends upon more than just the presence of peace or war. Most importantly, perhaps, it depends upon just how much money the president requests. At the margin, we expect that Congress will look more favorably upon smaller requests than larger ones. Wethereforecontrolforthel-oggedvalueofthepresident's proposal for each agency in each year. Congress's response to the president surely also depends upon the level of political support that he enjoys within its chambers. Presidents who confront congresses with large numbers of ideological or partisan supporters are likely to secure appropriations that more closely approximate their requests than presidents who face off against congresses dominated by the opposition party. Following Kiewiet and McCubbins (1985a, 1985b), we therefore control for the percent of House seats held by the president's party in each year. We also include three economic indicators: the average unemployment rate during the year when appropriation-sareproposedandset;thenationalgrowth rate since the previous year; and the total budget deficit fromtheprevious year.Onemightexpectthatpresidents receive greater popular support when the economy is doingwell,andfurther,thattheeconomymightdobetter in times of war due to increased government spending. By controlling for these three economic indicators, we preclude their ability to bias the effect of war on presidential bargaining success. All of our statistical models include fixed effects that account for all observable and unobservable timeinvariantcharacteristicsofindividualagenciesandpresidents.

23

Prior work shows that civilian killing in one period is de-pendent on the dynamics that allowed for killing in the previous period (Eck andHultman 2007). We includeAll One-Sided Violence(t-1), Rebel One-Sided Violence (t-1) ,and GovernmentOne-SidedViolence(t-

1)tocorrespond to their respectivedependent variables.Thesevariables shouldex-hibit a positive relationship with civilian killing. 3 We also expect that asbattlefieldviolence increases, hostilities spill over into the population (Downes 2008; Hultman 2007). We consequently control forAll BattleDeaths, Rebel Battle Deaths,andGovernment BattleDeaths,themonthlynum-ber of battlefield deaths incurred by all factions, rebels, and regime forces, respectively. We control for the source of conflict between the government and rebels to de-termine if it is significantly related to the targeting of noncombatants.Government Conflictis a dichotomous variable that uses the UCDP/PRIO delineation of civil wars fought over territorial (0) or government (1) con-trol.We also expect longer wars to offer greater incentives for factions to victimize civilians. As a war wears on and neither belligerent is able to subjugate the other, factions may turn to victimization to tip the balance.Conflict Du-rationis the number of months since the beginning of a conflict episode. Finally, a larger populationoffers greater opportunities for civilianmistreatment. We includePop-ulationtorecordeachwar country's yearlypopulationsize according to the Composite Index of National Material Capabilities data (Singer, Bremer, and Stuckey 1972)

24

The main controls in this study are similar to those used in previous studies of state repression (e.g., Davenport and Armstrong 2004; Poe and Tate 1994): regime type, economic development, population size, ongoing armed civil conflict (intrastate), and levels of dissent.We also use different indicators for some of these dimensions for robustness checks. Each control variable is discussed briefly below. First, it is a well-established finding in the literature that democracy is positively related to respect for personal integrity rights (e.g., Bueno deMesquita et al. 2005; Davenport 1995, 2007a; Hibbs 1973) and hence lower levels of state repression. Using the Polity database (Marshall and Jaggers 2005), Davenport and Armstrong (2004) established that there is an important threshold effect for when democracy matters for curbing repression: on and below a specific high value of democracy (7 on the Polity measure), there is no impact; above this value, there is a strong and negative influence in two distinct phases (one exerted at levels 8 and 9 as well as one exerted at level 10). Based on this, we follow the convention from Davenport and Armstrong (2004) and include a dummy for Polity scores 8–9 and a dummy for a Polity score of 10. These dummies have been found to be significantly different from the Polity scores below 8 in predicting repression and also statistically significantly different from each other (Davenport and Armstrong 2004). Other structural characteristics of states have also been found to predict repression, such as population size and development (e.g., Mitchell and McCormick 1998). We therefore control for total population size (natural log) and development measured by GDP per capita using data from PennWorld Tables, both measures from Urdal (2006). Previous studies have found that ongoing conflict activity may increase the risk that leaders will repress their citizens (Davenport and Armstrong 2004; Landman 2005; Poe 2004).We therefore include a dummy variable for ongoing intrastate armed conflict, using the Uppsala and PRIO Armed Conflict dataset (Gleditsch et al. 2002). A related finding is that dissent could increase repressive action (Davenport 1995; Hibbs 1973). Hence, we control for the annual number of antigovernment protests, riots, or strikes involving more than 100 persons from Banks (2002) followingWood (2008).Thismeasure varies between 0 and 46 in our sample.28 Ina data structure like the one used

here, there is likely no independence between all observations, andwe should expect that previous repression levels within a country matter for currently observed levels. Several studies find that a lagged dependent variable (LDV) of repression is highly significant (Davenport 2007a; Poe 2004), and we therefore run statistical models with LDVs, indicating the level of repression in the previous year (LDV = 2, LDV = 3, LDV = 4, LDV = 5).

25

Granted, randomassignment ofmunicipalities to different levels of BFP coverage would be the best way to assess any the program's electoral effects. However, there were no pilots, and the program was phased in rather quickly over the whole country. As a next best alternative, I employ a generalized propensity score (GPS) matching approach to attempt to hold fixed the connection between development and CCT coverage (Imai and van Dyk 2004) and focus on the independent contribution thatCCTsmake to incumbent candidate vote share by contrastingmunicipalities that differ with respect to CCT coverage but that have similar levels of development, size, political background, and other observable covariates. This matching procedure requires stratifying observations into similar groups. The important assumption here is that within groups, variations in coverage are as good as random. Following (Imai 2004), this was done by computing propensity scores—the treatment levels predicted by pretreatment covariates—and then partitioning the data into strata where all observations have similar propensity scores. The treatment effect is calculated within each strata by a simple linear regression of incumbent vote share on the treatment variable (CCT coverage), controlling for the propensity score itself and growth rates. GPS matching does not overcome the possibility of omitted variable bias. Nonetheless, there are significant advantages to this approach over simply estimating an OLS regression. At a minimum, the matching procedure ensures common support (i.e., that only similarmunicipalities are actually compared) and relaxes the assumption that causal effects are the same across all types of municipalities (i.e., it allows for heterogeneous treatment effects). These advantages are not trivial, as they make matching procedures less model dependent than OLS regressions. Moreover, the combination of both approaches, as employed here, allows for even greater robustness; regression serves as a correction for potentially faulty matching (Morgan and Winship 2007, 156), and matching serves as a protection from model misspecification (Ho et al. 2007). Matching will be a better statistical fix for nonrandom assignment the better the treatment can be predicted from observed covariates (Morgan and Winship 2007, 114). This is particularly relevant in the present case because there are strong theoretical and empirical reasons to believe the specific propensity score regression employed can ensure proper identification ofCCTeffects. Much of the strength of the empirical strategy relies on the inclusion of the government's official target of coverage for each municipality in the propensity score regression. This target was computed once by the government, based on the 2006 national household survey. Even though the target did not exist in 2002, it provides a reasonably neutral assessment of "need" in eachmunicipality because it is based on social indicators that move slowly over time. Besides the target, the most powerful predictor of treatment is precisely the HDI-M, to which I added squared and cubed terms to improve the fit of the propensity score regression at extreme levels of coverage, as well as the same control variables that were included in the OLS models. State fixed effects were added to account for variation

not captured by the substantive variables and only marginally improved the fit of the regressions. In the end, the regressions predicting treatment in each year had R2s of at least 0.8 and exhibited good fit across the full range of the treatment values. As a result, coverage effects are identified based on small deviations from the predicted levels of coverage, and it is reasonable to treat variations in CCT coverage within each strata as random.

26

We control for a range of other factors that might influence renewables growth. First, we include the previous share of renewables in electricity generation. Our model highlights the importance of endogenous growth in the renewable sector after exogenous shocks. We thus ex-pect that past generation capacity is a positive predictor of future growth. To smooth the data, we use a three-year average of past renewable shares (fromt?3tot?1). Per capita income and GDP growth are included be-cause wealth allows countries to invest in clean energy (Grossman and Krueger 1995). We also include the share of government expenditures of GDP. This accounts for the size of the government and thus proxies for public involvement in the national economy. Similarly, we control for the share of investment in a country's GDP to proxy for the country's general tendency to invest in pro-duction capacity. Trade openness, defined as the ratio of exports and imports to GDP, is also included because pre-vious research suggests that export industry creation is a key factor in clean-energy policy (Lewis and Wiser 2007; Lund 2009). All are measured in constant dollars from the Penn Tables (Heston, Summers, and Aten 2009). Finally, we also control for past de facto energy policy decisions by including the share of electricity produced by nuclear plants and hydro installations, as recorded by the United States Energy Information Administration. 28 An increase in oil price (our exogenous shock) may lead to an increase in such sources of electricity instead of the development of renewable sources. For instance, France's lack of renewable electricity production could be the result of reliance on nuclear power. Similarly, we include a measure for the generation share of conventional thermal electricity oil, natural gas, and coal used in power generation – to account for a government's political cost of going against the interests of the fossil fuel industry.

27

To deal with temporal dependence of recurrent conflict, the data are structured as a binary time-series cross-section, and I include a measure of the number of years since a civil war (onset or incidence) with cubic splines (Beck, Katz, and Tucker 1998). 26 Thus, the anal-ysis accounts for the amount of time since the last civil conflict broke out or was active. In the models presented here, I control for factors likely to influence both the extent of division in the SD movement and the onset of civil war. These include previous concessions to the movement, whether the host state is a democracy, and whether the movement has ge-ographically close kin. Concessions to SD movements suggest that the state is actively attempting tomanage the SDmovement's demands andmay decrease the chance of an armed challenge. In addition, concessions may satisfy some factions' demands and lead themto exit the dispute. Open competition in democratic states and the norm of respecting citizen demands could lead to SD movements havingmore factions.

Additionally, democracies are gen-erally expected to be less likely to experience civil war. 28 Movements with kin in an adjoining state may be more likely to form factions linked to these kin who seek to influence politics in their homeland. The existence of a neighboringstatewithethnickinmayalsoaffect the state's or movement's willingness to use force (Jenne 2006). 29 Table 3 reports the results of my analyses of civil war onset and civil war incidence.

28

Seven control variables capture country- and election-specific factors that may influence the probability of a democratic transition. First, it is likely that democratization occurs primarily in elections that deter-mine who will hold executive power. Because the stakes are higher in these contests, they are more likely to elicit higher levels of domestic mobilization for democracy. I include a dichotomous variable, main election,coded as a 1 for presidential elections in presidential (or mixed) systems and legislative elections in parliamentary systems (Simpser and Donno 2012). Second, I include a variable indicating whether theincumbent was running in the election, which is expected to decrease the chances of de-mocratization (Cheeseman 2010; Maltz 2007).15 It is also important to account for the country's previous experi-ence with elections. Lindberg (2006, 2009) argues that holding elections fosters institutional change and greater respect for civil liberties and that these changes cumulate over time. Relatedly, Bratton and Van deWalle (1997) ar-gue that the more elections held under authoritarianism, the greater the chances for democratization. I thus in-clude a variable that sums thenumber of previous elections held under a continuous authoritarian spell. 16 Two variables control for economic conditions: first avariableforGDP per capita,lagged one year. 17 High income is reliably associated with democracy, though its effect on democratic transitions is less clear. Przeworski et al. (2000, chap. 2), for example, find that dictatorships at high (but not the highest) levels of income are more likely to democratize. Second, I include a variable for GDP growth, measured as the percent change in a country's GDP from yeart-2tot-1.Ifitistruethatgood eco-nomic performance bolsters authoritarian regimes, the coefficient on this variable should benegative.

29

All estimations also include a set of controls. The first is Relative Wealth. Research suggests that poorer citizens are more likely to be persuadable by private goods offers than other voters (Dixit and Londregan 1996). For the very poor, immediate improvements of material conditions (even a small cash handout or a bag of rice) take priority over collective goods that come with credible commitment problems (Desposato 2007; Scott 1977). In new democracies, low trust in politicians has been found to exacerbate risk aversion of poorer voters, and low-information environments erect additional barriers for poor voters to enforce collective goods promises (e.g., Stokes 2000). Given the difficulties of reliably measuring absolute wealth, we adopted a measure used by the Afrobarometer. This measure asks respondents to evaluate their personal economic situation relative to other citizens and provides fairly reliable information about the relative economic condition of the respondent. The measure has five values and ranges from much poorer than average to much better off than average. Partisanship: Partisanship is expected to

make voters less likely to swing vote. It is measured as a dichotomous variable that is coded as 1 if the respondent indicated being active in a political party. Voted forWinning MP: Whether or not the respondent voted for the incumbent MP in the last election is used to condition the hypothesized effects of performance evaluations as indicated above, but we also expect it to have an independent effect. In the first instance, supporting a candidate who ultimately won may bias evaluations positively. In the second instance, one can expect that bad performance can push voters away from their candidate who won the previous elections. For challenger voters, poor performance may make them swing towards another candidate but certainly not the one in power who did a bad job of delivering goods. This is particularly important in Ghana, where MPs frequently report that constituents hold them responsible as agents of development (Lindberg 2010), and because all MPs have equally sized discretionary constituency development funds. Respondents who supported the candidate who won the 2004 elections are coded as 1 and all other respondents (voters for losers and nonvoters) as 0. Male: In the control for gender, men are coded 1 and womenare assigned 0.Research suggests thatwomenhave higher levels of risk aversion in political and economic activities (Eckel and Grossman 2008), and we thus expect them to exhibit more stable voting behavior. Age: The controls for the age of a subject-group's cohorts are 18–22, 23–35, 36–55, and 55 and older.16 We expect older individuals to have more entrenched voting habits and consequently are less likely to change them (Franklin 2004). Education: To capture formal education, the subjects' highest level of schooling is included. This ordinal variable with five levels ranges from no formal schooling to post-secondary/university education. Better-educated voters will have more adept reasoning skills and, all else equal, are more critical and evaluative. Hence, we expect clientelistic swing voting to be negatively associated with level of education while the intuition is that its relationship with policy swing voting should run in the opposite direction. Information: We expect that more informed voters will have greater confidence in their own political views and consequently are less likely to swing vote. An index capturing a subject's exposure to newsmedia based on the frequency with which she gets information from radio, newspapers, and television is included. Ethnicity: Two dummy variables derived from a question that asks respondents to identify their tribe,with 1 in both cases indicating being Ashanti or Ewe, respectively, were employed. Safe Havens: In safe havens, a single voter's ballot impact on electoral outcomes is close to 0, and thus the appeal of switching parties to obtain some end should be marginal.19 Constituency competitiveness is measured with a dummy variable where safe havens are coded as 1 when one party has won the last several elections with a margin of victory exceeding 20%. Other constituencies are coded as 0. Other measurement strategies and the robustness of the findings using other approaches towards this variable are detailed in online Appendix E.6.

30

As a robustness test, Models 1–6 in Table 1 omit all control variables, while Models 7–15 present the fully specified models. The results, though not identical, are quite similar, suggesting that the rela- tionships reported below are not artifacts of model spec- ification. Consequently, I proceed more confidently to interpreting the fully specified model.

31

Several theoretically relevant contextual- and individual-level controls were included in the analysis. First, prior research suggests that the economic and po-litical environment surrounding citizens may both exert distinct influences on their attitudes toward immigration (Campbell, Wong, and Citrin 2006). Data from the Bu-reau of Labor Statistics were utilized to create a measure of the unemployment rate in each respondent's county in 2005. The resulting variable is coded to range from low to high county unemployment. The second contextcommand ual variable captures the political climate surrounding respondents by measuring the percent of registered vot-ers in a respondent's county voting for Bush in the 2004 presidential election. 8 Turning to individual-level controls, all analyses included standard measures of educa-tion, income, gender (1=Male), age, citizenship status (dichotomous; 1=born in the United States), employ-ment status (1=Unemployed), pocketbook economic evaluations (1=experiencing financial distress), party identification (standard 7-point scale; 7=strong Repub-lican), and ideological self-identification (11-point scale; 11=very Conservative). Beyond these standard controls, several additional individual-level factors have been identified in the liter-ature for shaping general attitudes toward immigrants. Of these, prejudice (Citrin et al. 1997; Huddy and Sears 1995) and the strength of national identity (Sides and Citrin 2007; Sniderman, Hagendoorn, and Prior 2004) stand out as likely predictors of both immigration threat perceptions and policy preferences. All analyses include an 11-category measure of general negative affect toward Hispanics,Hispanic Affect, with high values indicating strong dislike for Hispanics. Given the present critique of cultural threat and the argument that residing in an acculturating context should serve as a tangible source of cultural threat that is separate from identity-oriented 8 These data were retrieved from the CNN 2004 Election Results cite listing vote results by county and state. For information, see http://www.cnn.com/ELECTION/2004/. 382 BENJAMIN J. NEWMAN concerns, controlling for national identity is essential. A measure of the strength ofNational Identitywas included in all analyses (1=strong national identity). In addition to prejudice and national identity, research has demon-strated that personality traits, such as authoritarianism, can influence threatperceptionsandopinionon immi-gration (Hetherington and Weiler 2009); all models in-clude a control forRight-Wing Authoritarianism.Last,in-tergroup contact theory suggests that having friends who are immigrants may reduce threat perceptions and in-crease support for permissive policy positions. To control for this possibility, all analyses included a dichotomous measure of whether or not respondents report having any close friends who are recent immigrants (1=has immi-grant friends). For ease of interpretation, all contextual-and individual-level independent variables were recoded to range from 0 to 1.

32

In order to test our hypotheses, we begin by estimating a series of OLS regressions where the dependent variable is a member's Legislative Effectiveness Score. Since our hypotheses concern the difference between women in the majority and minority parties, we include indicator variables for whether a legislator is Female, and either a Majority Party Female or a Minority Party Female. A Lagged Effectiveness Score is incorporated into the analysis

to control for the fact that members are expected to have consistent interest and innate abilities from one Congress to the next. Seniority and its squared value measure the number of terms that the member has served in Congress to capture the institutional influence that might be acquired by more senior members (and the squared value allows the seniority effect to taper off). While seniority is relevant to any investigation of legislative effectiveness, it is especially important to consider in the context of gender and politics, as it was not until the 109th Congress that women made up more than 15% of the House. Therefore, many female legislators have fewer years of experience than their male counterparts, which may be related to their abilities to be effective lawmakers. State Legislative Experience is a dummy variable that captures whether a member served in the state legislature prior to entering Congress. As Carroll points out, "many of the women who run for Congress have gained experience and visibility in state government before seeking federal office" (2004, 6). In fact, over 40% of the female representatives in the 107th Congress had served in their state's legislative body (Carroll 2004, 6), which one might expect would translate into increased effectiveness. Because state legislatures vary significantly in their professionalism, we also interact State Legislative Experience with an updated version of Squire's (1992) Legislative Professionalism measure to account for the possibility that members who served in more professional state legislatures will be more effective in Congress. Majority Party is a dummy variable for whether a member is in the majority party, which is thought to be important for policy advancement generally. Majority Party Leadership accounts for whether a member is among the leadership (majority party leader, deputy leader, whip, and deputy whip), with a similar variable included also for Minority Party Leadership. Speaker is a dummy variable for the Speaker of the House; Committee Chair captures whether a member is a chair of a standing committee; and Power Committee captures whether a member serves on the Rules, Appropriations, or Ways and Means Committees. All of these variables are particularly relevant as controls for this analysis, as female legislators have been generally less likely to attain these positions of influence, and we are interested in women's effectiveness when accounting for these institutional differences. Distance from Median captures the absolute distance between the member and the chamber median on the DW-NOMINATE ideological scale (Poole and Rosenthal 1997) to control for the possibility of more centrist members offering proposals that are more likely to find their way into law. Since previous research has demonstrated that female lawmakers are more liberal than their male counterparts, especially (until recently) when in the Republican Party (e.g., Burrell 1994; Frederick 2010; Swers 2005), this variable is particularly relevant to our study.8 Members' personal characteristics, including African American and Latino, are incorporated because they have been shown to be important in earlier studies of effectiveness (e.g., Griffin and Keane 2011; Rocca and Sanchez 2008). Size of Congressional Delegation within the member's state captures the possibility of natural coalitions among members from the same state. Vote Share and its square are included to allow for the possibility that members from safe seats can dedicate greater time and effort to internal legislative effectiveness rather than external electioneering and to allow this effect to be nonlinear.

References

Beckstead, Jason W. 2012. "Isolating and Examining Sources of Suppression and Multicollinearity in Multiple Linear Regression." *Multivariate Behavioral Research* 47(2):224–246.

Clarke, Kevin A. 2005. "The Phantom Menace: Omitted Variable Bias in Econometric Research." *Conflict Management and Peace Science* 22(4):341–352.

Clarke, Kevin A. 2009. "Return of the Phantom Menace: Omitted Variable Bias in Political Research." *Conflict Management and Peace Science* 26(1):46–66.

Kim, Yongnam. 2019. "The Causal Structure of Suppressor Variables." *Journal of Educational and Behavioral Statistics* p. 1076998619825679.

Lynam, Donald R., Rick H. Hoyle and Joseph P. Newman. 2006. "The Perils of Partialling: Cautionary Tales from Aggression and Psychopathy." *Assessment* 13(3):328–341.

Maassen, Gerard D. and Arnold B. Bakker. 2001. "Suppressor Variables in Path Models: Definitions and Interpretations." *Sociological Methods & Research* 30(2):241–270.

Middleton, Joel A, Marc A Scott, Ronli Diakow and Jennifer L Hill. 2016. "Bias Amplification and Bias Unmasking." *Political Analysis* 24(3):307–323.

Pandey, Shanta and William Elliott. 2010. "Suppressor Variables in Social Work Research: Ways to Identify in Multiple Regression Models." *Journal of the Society for Social Work and Research* 1(1):28–40.

Paulhus, Delroy L., Richard W. Robins, Kali H. Trzesniewski and Jessica L. Tracy. 2004. "Two Replicable Suppressor Situations in Personality Research." *Multivariate Behavioral Research* 39(2):303–328.

Steiner, Peter M. and Yongnam Kim. 2016. "The Mechanics of Omitted Variable Bias: Bias Amplification and Cancellation of Offsetting Biases." *Journal of Causal Inference* 4(2).