Pitfalls in the Development of Falsification Tests: An Illustration from the Recent Minimum Wage Literature

Clemens, Jeffrey

UC San Diego

14 June 2017

ESSPRI Working Paper Series Paper #20172

Pitfalls in the Development of Falsification Tests: An Illustration from the Recent Minimum Wage Literature

Economic Self-Sufficiency Policy Research Institute

Jeffrey Clemens, UC San Diego; ESSPRI

6-14-2017

Pitfalls in the Development of Falsification Tests:

An Illustration from the Recent Minimum Wage Literature

Jeffrey Clemens*

June 14, 2017

Abstract:

This paper examines a "falsification test" from the recent minimum wage literature. The analysis illustrates several pitfalls associated with developing and interpreting such exercises, which are increasingly common in applied empirical work. Clemens and Wither (2014) present evidence that minimum wage increases contributed to the magnitude of employment declines among low-skilled groups during the Great Recession. Zipperer (2016) presents regressions that he interprets as falsification tests for Clemens and Wither's baseline regression. He interprets his results as evidence that Clemens and Wither's estimates are biased. In this paper, I demonstrate that Zipperer's falsification tests are uninformative for their intended purpose. The properties of clustered robust standard errors do not carry over from Clemens and Wither's baseline specification (27 treatment states drawn from 50) to Zipperer's falsification tests (3 or 5 "placebo treatment" states drawn from 23). Confidence intervals calculated using a setting-appropriate permutation test extend well beyond the tests' point estimates. Further, I show that the sub-samples to which Zipperer's procedure assigns "placebo treatment status" were disproportionately affected by severe housing crises. His test's point estimates are highly sensitive to the exclusion of the most extreme housing crisis experiences from the sample. An inspection of data on the housing market, prime aged employment, overall unemployment rates, and aggregate income per capita reveals the test's premise that regional neighbors form reasonable counterfactuals to be incorrect in this setting.

---

*Clemens: University of California at San Diego, Jeffrey Clemens, Economics Department, 9500 Gilman Drive #0508, La Jolla, CA 92093-0508, USA. Telephone: 1-509-570-2690. E-mail: *clemens.jeffrey@gmail.com*. Significant portions of text draw heavily on the text of Clemens and Wither (2014), which this paper supplements. I thank David Neumark for comments. I am grateful to the Don and Sybil Harrington Foundation and the economics department at the University of Texas at Austin for their generous financial support and hospitality during the writing of this paper.

Falsification tests are an increasingly common element of applied empirical work. In simple Google Scholar searches, for example, the search terms "falsification test" and "economics" appear jointly in only 28 entries from 1990 to 1999. In 2016 alone, the phrases appear together in 886 entries, up from 687 in 2015 and 183 in 2010.[1]

The intent of falsification exercises is to provide a test that is informative regarding the validity of an underlying estimate and/or estimation strategy. In practice, however, it can be difficult to evaluate whether such tests have been implemented on a systematic or ad hoc basis, whether appropriate inference procedures have been applied, and whether the test's point estimates are prone to bias. I use a falsification test from the recent minimum wage literature to illustrate these issues.

During the Great Recession and subsequent recovery, employment declined dramatically among individuals with low levels of experience and education. The analyses in Clemens and Wither (2014) and Clemens (2015) explore the extent to which the federal minimum wage, which rose from $5.15 to $7.25 over this time period, contributed to these declines. These analyses make use of the fact that the $7.25 federal minimum wage was differentially binding across states. Using standard difference-in-differences, triple-difference, and matching estimators, Clemens and Wither (2014) and Clemens (2015) estimate that this period's minimum wage increases explain a non-trivial fraction of the decline in low-skilled groups' employment.[2]

In a recent working paper, Zipperer (2016) critiques selected portions of a September 2016 revision of Clemens and Wither (2014).[3] Zipperer summarizes his findings by

[1]These searches were conducted on March 19, 2017. The total numbers of entries containing the phrase "economics" were lower in 2016 and 2015 than in 2010, suggesting that the propensity for economics articles to use the phrase "falsification test" has increased more dramatically than the raw counts imply.

[2]A September 2016 revision of Clemens and Wither (2014) consolidates these empirical analyses. This revision can be found at the following link: `http://econweb.ucsd.edu/~j1clemens/pdfs/ClemensWitherMinimumWageGreatRecession.pdf`.

[3]The December 2016 version of the critique, which is modestly updated from a November 2016 version, can be found at the following link: `http://cdn.equitablegrowth.org/wp-content/uploads/`

writing: "I find that the authors' baseline results are not robust to sectoral or geographic controls, which reduce the magnitude of the baseline point estimates by 35.6 to 62.7 percent. Moreover, their research design fails a placebo-based falsification test." My focus in this paper is on the properties of Zipperer's "placebo-based falsification test."[4] I show that this exercise falls far short of what Zipperer claims it delivers. I highlight several distinct issues that, in addition to being relevant to this particular setting, may be of interest to researchers tasked with evaluating or developing falsification tests in other settings.

A first set of issues involves statistical inference. Falsification exercises may have different inference properties than the baseline they are being used to evaluate. In the present case, this is relevant for two reasons. First, the falsification test is conducted on a subset of the baseline analysis sample. Second, the test's sample is less evenly divided into "treatment" and "control" units than the baseline analysis sample. Further, because the exercise's development comes with many degrees of freedom, multiple hypothesis testing concerns should be kept in mind.

I demonstrate that Zipperer's falsification tests are uninformative for their intended purpose. Zipperer implicitly assumes that the properties of clustered robust standard errors carry over from Clemens and Wither's baseline specification to his falsification test. Simulations from Bertrand, Duflo, and Mullainathan (2004) and Cameron, Gelbach, and Miller (2008) suggest that this assumption is unlikely to hold. Clemens and Wither's baseline specification involves a sample in which there are 27 treatment states drawn from 50 total states. By contrast, Zipperer's falsification tests assign placebo treatment status within a sample consisting of Clemens and Wither's 23 control states. One test assigns placebo treatment status to 3 of these 23 states; the other assigns placebo treat-

---

2016/12/02155549/120616-WP-comments-on-clemens-and-wither.pdf.

[4] I save a fuller analysis of Zipperer's preferred specification for Clemens (2017), which connects several applied econometric issues into a longer running debate within the minimum wage literature.

ment status to 5 of these 23 states. I show that confidence intervals constructed using a setting-appropriate permutation test (Abadie, Diamond, and Hainmueller, 2012; Imbens and Rosenbaum, 2005) are, on average, more than twice as wide as the confidence intervals Zipperer reports. They extend well beyond both the point estimate from Clemens and Wither's baseline and the point estimates from the falsification exercises. In contrast, the confidence interval associated with Clemens and Wither's baseline changes little when calculated using permutation test methods rather than clustered robust standard errors.

This initial point, namely that Zipperer's test is uninformative, is important to bear in mind throughout. My subsequent empirical analysis reveals that the point estimates from Zipperer's falsification tests are quite sensitive. The observed sensitivity (e.g., point estimates changing sign and increasing in magnitude) would be surprising if the standard errors he reports were correct. It is far less surprising in light of the confidence intervals I estimate using the permutation test approach.

A second set of issues involve sources of bias. My general point is that falsification tests are sufficiently distinctive exercises that their development and evaluation requires attention to test-specific sources of bias. A first issue is that threats to identification can be shaped by the algorithm a falsification test uses in its assignment of "placebo treatment status" across sampled units.[5] As detailed below, this has high relevance in the current setting. A second issue is that the appropriate null hypothesis for a falsification

[5]A separate issue applies in settings where falsification tests involve assessing whether a treatment is correlated with changes in an outcome that it should not affect. In this case, it is important to consider what sources of bias might affect the new outcome under analysis. In analyses of the health benefits of health insurance, for example, deaths from "internal causes" are sometimes the outcome of interest while deaths from "external causes," which include accidental deaths, are are potentially used as a falsification test. In this case, the baseline analysis of disease-related deaths could safely ignore correlations between the policy change of interest and policy changes involving traffic safety and gun laws. Traffic safety and gun laws may be quite relevant, however, to the validity of the falsification test. Changes in such laws may have effects on accidental deaths that, if not taken into account, could bias the falsification test in one direction or another.

test may not be 0. In many settings, including the current setting, the appropriate null may involve a muted effect or an effect of opposite sign.[6]

I show that the states to which Zipperer's procedure assigns "placebo treatment status" were disproportionately affected by severe housing crises. A majority of the individuals in the placebo treatment sample in Zipperer's "region-based" exercise reside in Florida. Similarly, a majority of the placebo treatment sample in his "division-based" exercise reside in either Florida or Arizona. Readers familiar with the geography of the housing crisis will recognize this distribution of the placebo treatment sample across states to be a likely source of bias. I show that the point estimates from Zipperer's tests are highly sensitive to these states' inclusion in the sample.

The premise underlying Zipperer's exercise is that regional neighbors provide reasonable counterfactuals for one another.[7] The plausibility of this premise is readily investigated by examining the evolution of macroeconomic covariates across the states that are proposed as one another's counterfactuals. An inspection of state level data on the housing market, prime aged employment, overall unemployment rates, and aggregate income per capita reveals the test's premise to be incorrect.

Finally, I investigate whether the specifications Zipperer advances as being preferred to the Clemens and Wither baseline satisfy his own falsification test. They do not. The falsification test would, if taken at face value, support the conclusion that Zipperer's preferred specifications are biased towards finding positive effects of this period's minimum wage increases on employment.

I conclude by emphasizing that Zipperer's falsification test should be taken as evi-

[6]As Zipperer reports in his table 3, placebo treatment is modestly positively correlated with actual minimum wage increases. The appropriate null hypothesis for his falsification test is thus not 0 if viewed as a test for whether Clemens and Wither's baseline estimate is unbiased.

[7]The exercise is thus relevant to the broader minimum wage literature because this is also the premise underlying Allegretto, Dube, Reich, and Zipperer's (2017) critique of work by Neumark and Wascher (2006), Neumark, Salas, and Wascher (2014b), and Meer and West (2016).

dence against *neither* his preferred specification nor the baseline estimate from Clemens and Wither. As discussed above, the test is uninformative for its intended purpose. Further, the biases to which the test is exposed are quite clear from the data presented in section 3. The test's sensitivity, as explored in section 5, likely reflects both of these issues.

The remainder of this paper proceeds as follows. Section 1 provides background on the research setting. Section 2 describes the falsification test Zipperer advances. Section 3 presents macroeconomic data that are informative for assessing the premise underlying the falsification test. Section 4 presents my evaluation of the falsification test's inference properties. Section 5 presents my analysis of the test's sensitivity to its inclusion of states with extreme housing crisis experiences. Section 5 also explores how Zipperer's preferred specification performs under his own falsification test. Section 6 concludes.

1 Background Regarding the Research Setting

1.1 Variation Generated by the Increase in the Federal Minimum Wage

This section provides an abbreviated description of the minimum wage changes analyzed in Clemens and Wither (2014). Legislation passed in May 2007 caused the federal minimum wage to rise from $5.15 to $7.25 in 3 increments of 70 cents. The increments were implemented in July 2007, July 2008, and July 2009. The analysis makes use of the fact that, to a first approximation, states were either fully bound by the final 70 cent increment or unbound by that increment. This motivates a division of the states into a treatment group of "bound" states and a control group of "unbound" states. The division of states and the evolution of effective minimum wage rates can be found in figures 1 and 2 respectively, both of which appear in Clemens and Wither (2014).

1.2 Estimation Strategies

Clemens and Wither (2014) use wage data from the 2008 panel of the Survey of Income and Program Participation (SIPP) to identify the least skilled workers within these treatment and control groups.[8] The most basic estimation framework in the original analysis is a difference-in-differences framework. The analysis compares the employment and income trajectories of low-skilled individuals in bound states to low-skilled individuals in unbound states. That is, it follows the trajectories of individuals who were employed at low wage rates during the 2008 SIPP panel's first year, which corresponds with the 12 months preceding the federal minimum wage's July 2009 rise from \$6.55 to \$7.25.[9]

The difference-in-differences analysis takes the following form:

$$Y_{i,s,t} = \sum_{p(t)\neq 0} \beta_{p(t)} \text{Bound}_s \times \text{Period}_{p(t)} + \alpha_{1_s} \text{State}_s + \alpha_{2_t} \text{Time}_t + \alpha_{3_i} \text{Individual}_i + \mathbf{X_{s,t}}\gamma + \varepsilon_{i,s,t}. \quad (1)$$

Equation (1) includes controls for state, State_s, time, Time_t, and individual-specific, Individual_i, fixed effects. The vector $\mathbf{X_{s,t}}$ contains time varying controls for each state's macroeconomic conditions.

Graphical evidence presented in Clemens and Wither (2014) motivates the dynamics for which the specification allows. The analysis characterizes May to July 2009 as a

[8]The primary analysis sample consists of individuals whose average wage rates between August 2008 and July 2009 were less than \$7.50. In response to both Zipperer and an anonymous referee, Clemens and Wither (2014) have considered the robustness of their estimates to alternative sample construction procedures. A first alternative approach analyzed samples selected based on the distance between an individual's baseline wage rates and his or her state's minimum wage. A second alternative approach analyzed samples selected based on individuals' percentiles within bound and unbound states' wage distributions. These results are presented elsewhere.

[9]Clemens (2016) uses the same estimation framework to analyze a variety of program participation and budget related outcomes.

"Transition" period, prior months as the baseline, August 2009 through July 2010 as period "Post 1," and all subsequent months as period "Post 2."

In this setting, the most obvious threat to econometric identification is the possibility that the forces underlying the Great Recession were either more or less severe in the "bound" group than in the "unbound" group. To gauge the relevant biases, Clemens and Wither (2014) report a set of macroeconomic time series that include two labor market aggregates, overall economic output, and a proxy for housing market conditions. I reproduce and further discuss these macroeconomic time series in section 3. The key fact is that all four macroeconomic covariates reveal the underlying recession to have been more severe in the states that comprise the control group than in the states that comprise the treatment group.

The first portion of Zipperer's analysis involves augmenting equation (1) with interactions between time effects and variables that describe states' industry shares in 2005 and 2006:

$$\begin{aligned} Y_{i,s,t} = & \sum_{p(t)\neq 0} \beta_{p(t)} \text{Bound}_s \times \text{Period}_{p(t)} + \alpha_{1_s} \text{State}_s \\ & + \alpha_{2_t} \text{Time}_t + \alpha_{3_i} \text{Individual}_i + \mathbf{X_{s,t}}\gamma \\ & + \alpha_{3_i} \text{Ind. Share}_s^{\text{05-06}} \times \text{Time}_t + \varepsilon_{i,s,t}. \end{aligned} \tag{2}$$

At the conclusion of this paper's analysis, I apply Zipperer's falsification tests to his specification in which Ind. $\text{Share}_s^{\text{05-06}}$ represents a vector of industry shares across a set of 10 NAICS supersectors. This specification underlies two of Zipperer's graphical presentations, as well as some of his strongest quantitative claims.[10] A more dedicated

[10]For reasons that will become apparent below, the falsification test cannot be applied to Zipperer's specifications that contain time varying geographic fixed effects. This is because these control sets are are perfectly collinear with the variation in "placebo treatment" status.

assessment of the relative merits of specifications (2) and (1) can be found in Clemens (2017).

2 Understanding the "Falsification Test"

This section lays out the exercise Zipperer presents as a falsification test for Clemens and Wither's baseline specification. To the extent possible, I quote from Zipperer's working paper to describe its mechanics. After describing the exercise, I provide more detailed commentary regarding its intent, the premise underlying its validity, and the information that is relevant for assessing that premise.

The sample analyzed in Zipperer's falsification test is a subset of the low-wage sample analyzed in Clemens and Wither's baseline specification. Specifically, it consists of the subset of these individuals who reside in the control group, meaning the 23 states that were not fully bound by this period's federal minimum wage changes. Within that set of states, Zipperer assigns placebo treatment status to unbound states in predominantly bound-state regions. As Zipperer describes the procedure, "I focus on the unbound states, which by construction do not receive treatment, and assign them a placebo increase in the minimum wage if they reside in a Census division or region where a majority of the population lives in actually bound states."

On a sample restricted to the 23 unbound states, Zipperer then estimates

$$\begin{aligned} Y_{i,s,t} = &\sum_{p(t)\neq 0} \beta_{p(t)} \text{Placebo Bound}_{r(s)} \times \text{Period}_{p(t)} + \alpha_{1_s} \text{State}_s \\ &+ \alpha_{2_t} \text{Time}_t + \alpha_{3_i} \text{Individual}_i + \mathbf{X_{s,t}}\gamma + \varepsilon_{i,s,t}, \end{aligned} \tag{3}$$

where $r(s)$ indexes regions. Zipperer goes on to observe that "The South is the only

region-bound region." This refers to the fact that, for the region-based test, placebo treatment status is assigned to "unbound" states in the south. The test's control group thus consists of unbound states in census regions other than the south. Zipperer also conducts a version of the test that uses census divisions rather than regions. He writes that "In this case, the South Atlantic and Mountain regions are the division-bound areas containing unbound states."

Zipperer's interpretation of his tests' results is that "a negative estimate for [$\beta_{p(t)}$ from equation (3)] suggests that CW's main specification conflates minimum wage effects with omitted regional shocks." Negative estimates are precisely what Zipperer finds, as shown in table 1, which reproduces Zipperer's (2016) table 3. My replication appears in subsequent tables.

An appreciation of the logic underlying Zipperer's interpretation of his falsification test may require a clearer statement of the exercise's key assumption. The key assumption is that minimum wage changes are as good as randomly assigned within the regions and divisions Zipperer designates as "region bound" and "division bound." Put differently, the key assumption is that the bound and unbound states within these regions experienced economic shocks of similar magnitude. If true, Zipperer would be correct to interpret a non-zero estimate of $\beta_{p(t)}$ as reflecting differences in economic conditions that would also bias the estimates from Clemens and Wither's baseline.

Interestingly, a more broadly applied version of this assumption is at the heart of the debate over analyses of a three-decade panel of state and federal minimum wage changes.[11] The relevant question involves the issue of whether variation that extends

[11]Neumark, Salas, and Wascher (2014a) estimate negative employment effects using variations in minimum wage policy that extend both across and within census regions. Allegretto, Dube, Reich, and Zipperer (2017) argue that unbiased estimates can only be obtained by restricting attention to within-region variations in minimum wage policy. This is implemented in practice by adding time-varying geographic fixed effects to the regression specification. The analysis in Dube, Lester, and Reich (2010), which criticizes a larger body of work underlying Neumark and Wascher's (2006) book, includes a placebo test that follows the same logic as the placebo test in Zipperer (2016).

both across and within regions is more or less prone to bias than variations that occur exclusively within regions. As Neumark, Salas, and Wascher (2014a) and Neumark and Wascher (2017) point out, a useful analogy can be made between within-region variation in minimum wage policy and "within-family" approaches to estimating the returns to education. The key point is that stand applied econometric considerations make clear that "within" comparisons can plausibly generate estimates that are either more or less biased than comparisons that extend both "across" and "within" families or regions.

3 What Potential Biases Are Implied by Macroeconomic Indicators?

This section presents evidence on the evolution of several broad macroeconomic indicators that describe cross-state variations in the severity of the forces underlying the Great Recession. I emphasize two sets of comparisons. The first comparison involves the full sets of "bound" and "unbound" states analyzed by Clemens and Wither's baseline. These comparisons are relevant for gauging the biases to which Clemens and Wither's estimation strategy is exposed. I next compare the states to which Zipperer assigns placebo treatment status to their regional neighbors. These are the comparisons that are relevant for assessing whether his "falsification tests" are informative regarding the validity of the Clemens and Wither baseline.[12]

[12]Recall from above that Zipperer's key assumption is that the placebo treatment states experienced economic shocks that were similar to those experienced by their neighbors, which are among the actual treatment states in Clemens and Wither's analysis. The comparisons made below are thus precisely the comparisons of interest for assessing that assumption.

3.1 Comparing the Housing Crises in Treatment and Control States

Figure 3 presents times series for a set of macroeconomic indicators presented originally in Clemens and Wither (2014). Two of the series, the state unemployment rate and employment to population ratio, are comprehensive employment indicators from the Bureau of Labor Statistics (BLS). A third series, taken from the Bureau of Economic Analysis (BEA), describes total state economic output. The final indicator, a median house price index from the Federal Housing Finance Agency (FHFA), proxies for the housing market component of states' recessionary environments.

Clemens and Wither (2014) use the data in figure 3 to provide evidence that the Great Recession was more severe in their control group than in their treatment group. From 2006 through 2012, aggregate income declined by roughly $1,000 per capita more in the unbound states than in the bound states. The aggregate unemployment rate increased more than 1 percentage point more in the unbound states than in the bound states and the employment to population ratio declined by roughly 1 percentage point more. Finally, median house prices declined by roughly $80,000 more in the unbound states than in the bound states. The macroeconomic indicators in figure 3 thus point uniformly to the conclusion that the Great Recession was more severe in the states that comprise Clemens and Wither's (2014) control group than in the states that comprise their treatment group. Failure to control for this difference would tend to bias estimated employment impacts towards 0.

3.2 Examining the Severity of the Housing Crisis in the Placebo Treatment States

I now focus more narrowly on the macroeconomic developments that are relevant for assessing Zipperer's falsification test. As explained above, the test's key assumption

is that minimum wage changes are as good as randomly assigned within the census regions and divisions Zipperer designates as "region bound" and "division bound." That is, his interpretation requires that the states he codes as "placebo treatment" states are appropriate counterfactuals for their regional neighbors. The assumption can be summarized as requiring that the actual treatment states and placebo treatment states experienced similar macroeconomic shocks.

Zipperer reports that the South is the source of placebo treatment states in his region-based exercise and that the South Atlantic and Mountain West are the sources of placebo treatment states in his division-based exercise. While he does not list the states to which the "placebo treatment" designation is applied, they can be inferred by using figure 1 to identify the "unbound" states in these regions. In the South and narrower South Atlantic, the unbound states are Florida, West Virginia, and the District of Columbia.[13] In the Mountain West, they are Arizona and Colorado. In the region-based exercise, more than 80 percent of the relevant population thus comes from Florida. In the division-based exercise, more than 75 percent of the relevant population thus comes from Florida or Arizona. Readers familiar with the geography of the housing crisis will recognize this to be problematic, as Florida and Arizona experienced two of the most extreme housing crises around the country.

Tables 2 and 3 present data on macroeconomic conditions that compare the Great Recession's severity in the placebo treatment states to the states for which they are being used as counterfactuals. Table 2 reports data underlying Zipperer's region-based comparisons while Table 3 reports data underlying Zipperer's division-based comparisons. I present changes over two time periods of interest. In both tables, Panel A presents changes for which the base period precedes both the crisis and the federal minimum

[13]In principle this group also includes Delaware. In practice, however, the low-wage SIPP simple contains no Delaware residents.

wage legislation under analysis. Panel B presents changes for which the base period is the base period associated with the SIPP regression analysis, namely August 2008 through April 2009.[14] In both panels, changes are calculated from the relevant base period through the "post" period from the regression analysis, namely August 2009 through the end of 2012.

The data speak with near uniformity. The Great Recession was significantly more severe in the states to which Zipperer assigns placebo treatment status than in the states for which they are being used as counterfactuals. The comparison between the placebo treatment states and other states in the South is particularly striking. Consider the changes described in panel A. On average, the placebo treatment individuals were in states in which median house prices declined by $156,000, prime age employment declined by 6.5 percentage points, the overall unemployment rate rose by 6 percentage points, and personal income per capita declined by $3,800. In the remainder of the South, by contrast, median house prices declined by $14,000, prime age employment declined by 3.8 percentage points, the overall unemployment rate rose by 3.7 percentage points, and personal income per capita declined by $500. The difference in the severity of the Great Recession was thus quite significant. Differences over the shorter time horizon examined in panel B are more moderate but still substantial. Table 3 reveals that the differences relevant to the division-based exercises are also quite substantial.

Comparisons between the placebo treatment states and their regional neighbors are central to the falsification test's validity. The test's presumption is that the regional neighbors of the placebo treatment states experienced comparable economic shocks. Tables 2 and 3 reveal this presumption to be incorrect.

Tables 2 and 3 also present data on the macroeconomic experiences of the placebo

[14]Recall that, motivated by dynamics observed in both the wage and employment data, the regression codes May 2009 through July 2009 as a "Transition" period.

control states. Differences between the experiences of the placebo treatment and placebo control states can be found in column 5. The differences are, once again, quite substantial. This reveals that estimates from the falsification test are likely to be biased in the direction of finding a negative relationship between placebo treatment status and the employment of low-skilled individuals.

4 Inference within the Falsification Test

This section presents evidence on the confidence intervals associated with Zipperer's falsification test. Zipperer implicitly assumes that the properties of clustered robust standard errors carry over from Clemens and Wither's baseline specification to his falsification exercise. Simulations from Bertrand, Duflo, and Mullainathan (2004) and Cameron, Gelbach, and Miller (2008) suggest that this assumption is unlikely to hold. Clemens and Wither's baseline specification involves a sample in which there are 27 treatment states drawn from 50 total states. By contrast, Zipperer's falsification tests assign placebo treatment status within a sample consisting of Clemens and Wither's 23 control states. The region-based test assigns placebo treatment status to 3 states, while the division-based test assigns placebo treatment status to 5 states.

I estimate confidence intervals for both of Zipperer's falsification tests, as well as for the Clemens and Wither baseline, using both clustered robust standard errors and permutation test methods (Abadie, Diamond, and Hainmueller, 2012; Imbens and Rosenbaum, 2005). The permutation test procedure involves estimating a distribution of "placebo treatment effects," the dispersion of which provides evidence on the statistical uncertainty underlying the point estimates of interest. The 95 percent confidence interval implied by the permutation test is simply the range of placebo treatment estimates such that 2.5 percent of the estimates fall both above and below the interval.

The exercise preserves the underlying estimator's structure with respect to the number of units from which treatment status is drawn and the number of units to which treatment status is applied. For Zipperer's region-based exercise, for example, each iteration of the procedure assigns placebo treatment status to 3 of the 23 states on which the estimator was implemented. For the division-based exercise, each iteration assigns placebo treatment status to 5 of the 23 states on which the estimator was implemented. For the Clemens and Wither baseline, each iteration assigns placebo treatment status to 27 of the 50 states on which the original estimator was implemented. In each case, the distribution of placebo treatment effects is the result of 500 iterations of the procedure.

Figures 4 and 5 present the resulting distributions. Table 4 describes the resulting confidence intervals and compares them with the confidence intervals associated with clustered robust standard error estimation. For the falsification tests, the confidence intervals calculated using the permutation test approach are, on average, more than twice as wide as the confidence intervals Zipperer reports. They extend well beyond both the point estimate from Clemens and Wither's baseline and the point estimates from the falsification exercises. This reveals the falsification tests to be uninformative for their intended purpose. In contrast, the confidence interval associated with Clemens and Wither's baseline changes little when calculated using permutation test methods rather than clustered robust standard errors.

An additional inference- and interpretation-related issue is that the significance tests Zipperer conducts are not for the appropriate hypothesis test. As reported in Zipperer's table 3, "placebo treatment" status is modestly positively associated with actual minimum wage changes. That is, rather than being a "placebo" test it is a test involving a "treatment" just over 1/6th the size of the baseline treatment.[15] Clemens and Wither's

[15]Zipperer's Table A.1 reveals that the treatment increases to nearly 1/3 the size of the baseline treatment when he uses the full-sample time horizon rather than ending the sample in December 2011.

(2014) baseline estimate would, given the placebo treatment's correlation with minimum wage changes, imply an employment estimate of -0.012. This, rather than 0, is the appropriate null if Zipperer's test is to be interpreted as a test for whether Clemens and Wither's baseline estimate is unbiased.

5 The Sensitivity of the Falsification Test

Tables 5 and 6 present evidence on the sensitivity of Zipperer's falsification tests. The tables differ modestly with respect to the samples analyzed. Modest data processing differences led to a slight difference between Clemens and Wither's original analysis and Zipperer's near replication. These difference translate into differences in the samples on which the falsification test is estimated. The samples underlying the estimates in table 5 correspond with Zipperer's data processing procedure, while the samples underlying the estimates in table 6 correspond with Clemens and Wither's original analysis sample.[16] In both tables, the estimates in panel A apply sample weights while the estimates in panel B do not.

I explore the tests' sensitivity to taking steps to reduce the biases associated with the comparisons on which they rely. It should be borne in mind that, as shown in the previous section, all of the falsification test estimates are statistically uninformative for their intended purpose. The estimates in tables 5 and 6 are meant primarily to illustrate that the test's point estimates exhibit the sensitivity the permutation test implies.

My most basic approach to demonstrating the falsification tests' sensitivity is to estimate the tests on samples that exclude states that experienced relatively extreme housing

[16]Note that although I have perfectly replicated Zipperer's near replication of the Clemens and Wither baseline, there is a very slight difference between the falsification test he reports and my replication of his test using his data processing choices. The sample differs by 4 observations and the point estimates differ very modestly. Note that an initial indication of the test's sensitivity is that the data processing differences between the Clemens and Wither baseline and Zipperer's near replication have much larger effects on point estimates from the falsification test than on point estimates from the baseline specification.

crises. My first approach is to simply exclude Florida and Arizona from the sample. A comparison of column 2 to column 1 and of column 6 to column 5 reveals that the falsification test is rather wildly sensitive to these states' inclusion. For the region-based exercise, "post-1" point estimates shift from being moderately large and negative to being even larger and positive. Medium-run estimates become positive in most cases and are statistically indistinguishable from 0 (even using clustered robust standard errors) in all cases. For the division-based exercise, the "post-2" estimate moves moderately while "post-1" estimates change from being moderately large and negative to being moderately large and positive.

Columns 3 and 7 reveal that similar results are obtained when sample inclusion is based on a more systematic matching procedure. Specifically, I restrict the sample to the states that satisfy a matching criterion developed and implemented for a September 2016 revision of Clemens and Wither's original analysis. The procedure involves matching "bound" and "unbound" states on the basis of the severity of their median house price declines, then dropping from the sample all treatment and control states for which a match within $20,000 could not be found.[17]

Finally, the estimates in columns 4 and 8 show that the test is somewhat sensitive to increasing the flexibility of the controls for variations in states' macroeconomic conditions. Specifically, I include controls for the housing price index, unemployment rate, overall employment rate, and personal income per capita, each interacted with time period fixed effects. This specification uniformly shifts the falsification tests' "post 2" point estimates towards 0. The estimates from the region-based exercise shift quite close to 0. The estimates from the division-based exercise shift more modestly towards 0.

In a final exercise, I consider the performance of Zipperer's preferred specification

[17]The specific procedure applied was nearest neighbor matching without replacement. This approach to sample selection is sometimes called the "caliper" approach to matching.

when subjected to his own falsification test. The results of this exercise appear in table 7. Column 1 replicates the Clemens and Wither baseline and column 2 shows that I have almost perfectly replicated Zipperer's "industry-by-time controls" specification.[18] Columns 3 and 4 show that Zipperer's falsification test generates economically large and positive coefficients when applied to his preferred specification. If these estimates are interpreted in the same way Zipperer interprets his application of the falsification tests to the Clemens and Wither's baseline, they would be taken as evidence of substantial bias. More specifically, they suggest that Zipperer's preferred specification is biased against finding evidence of negative employment effects of the minimum wage changes under analysis.

6 Conclusion

This paper examines a falsification test from the recent minimum wage literature. The test is presented by Zipperer (2016), who interprets its results as evidence that estimates from Clemens and Wither (2014) are biased. My examination of Zipperer's test considers both its inference properties and the potential biases to which it is exposed.

Section 4 shows that the standard errors Zipperer (2016) reports provide a misleading impression of precision. The mistaken assumption is that the falsification test's inference properties are the same as those of the Clemens and Wither (2014) baseline against which it is advanced as a test. Section 4 shows that this assumption is incorrect and that the test is uninformative for its intended purpose.

Section 3 shows that the falsification test is subject to substantial bias. Data on aggregate employment, economic output, and the housing market show that the test relies

[18]The change in the point estimates between this specification and the Clemens and Wither baseline is almost exactly the same as the change in the point estimates between Zipperer's industry-by-time specification and his near replication of the Clemens and Wither baseline.

on comparisons between states across which the severity of the Great Recession differed dramatically. The test's premise that regional neighbors form reasonable counterfactuals is, in this setting, incorrect.

I conclude by emphasizing a more general problem with the increasingly prevalent use of "falsification tests" in applied econometric analyses. The frequently ad hoc nature of falsification tests renders them less informative than their statistical properties may appear to imply. Sufficient specification searching can inevitably generate a "test" that a) sounds *ex ante* reasonable, b) appears to discredit the hypothesis or specification the researcher is attempting to discredit, and/or c) appears to support the hypothesis or specification the researcher is attempting to advance. The development of such exercises involves more artistic license than is widely acknowledged. This paper's purpose is, in part, to draw attention to the fact that "placebo" and "falsification" tests can fall far short of the scientific authenticity implied by the associated terminology. The increasingly pervasive use of this terminology is worth reconsidering.

References

Abadie, A., A. Diamond, and J. Hainmueller (2012): "Synthetic control methods for comparative case studies: Estimating the effect of Californias tobacco control program," *Journal of the American statistical Association*.

Allegretto, S., A. Dube, M. Reich, and B. Zipperer (2017): "Credible Research Designs for Minimum Wage Studies," *ILR Review*, 70(3), 559–592.

Bertrand, M., E. Duflo, and S. Mullainathan (2004): "How Much Should We Trust Differences-in-Differences Estimates?," *Quarterly Journal of Economics*, 119(1).

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008): "Bootstrap-based improvements for inference with clustered errors," *The Review of Economics and Statistics*, 90(3), 414–427.

Clemens, J. (2015): "The Minimum Wage and the Great Recession: Evidence from the Current Population Survey," *NBER Working Paper 21830*.

——— (2016): "Redistribution through Minimum Wage Regulation: An Analysis of Program Linkages and Budgetary Spillovers," in *Tax Policy and the Economy, Volume 30*, ed. by J. Brown. University of Chicago Press.

——— (2017): "The Minimum Wage and the Great Recession: A Response to Zipperer and Recapitulation of the Evidence," *Mimeo*.

Clemens, J., and M. Wither (2014): "The Minimum Wage and the Great Recession: Evidence of Effects on the Employment and Income Trajectories of Low-Skilled Workers," *NBER Working Paper 20724*.

Dube, A., T. W. Lester, and M. Reich (2010): "Minimum wage effects across state borders: Estimates using contiguous counties," *The review of economics and statistics*, 92(4), 945–964.

Imbens, G. W., and P. R. Rosenbaum (2005): "Robust, accurate confidence intervals with a weak instrument: quarter of birth and education," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168(1), 109–126.

Meer, J., and J. West (2016): "Effects of the Minimum Wage on Employment Dynamics.," *Journal of Human Resources*, 51(2).

Neumark, D., J. I. Salas, and W. Wascher (2014a): "More on recent evidence on the effects of minimum wages in the United States," *IZA Journal of Labor policy*, 3(1), 24.

——— (2014b): "Revisiting the Minimum WageEmployment Debate: Throwing Out the Baby with the Bathwater?," *Industrial & Labor Relations Review*, 67(3 suppl), 608–648.

Neumark, D., and W. Wascher (2006): "Minimum wages and employment: A review of evidence from the new minimum wage research," *NBER Working Paper 12663*.

——— (2017): "Reply to Credible Research Designs for Minimum Wage Studies," *ILR Review*, p. 0019793917698429.

Zipperer, B. (2016): "Did the minimum wage or the Great Recession reduce low-wage employment? Comments on Clemens and Wither (2016)," Discussion paper.

Figures and Tables

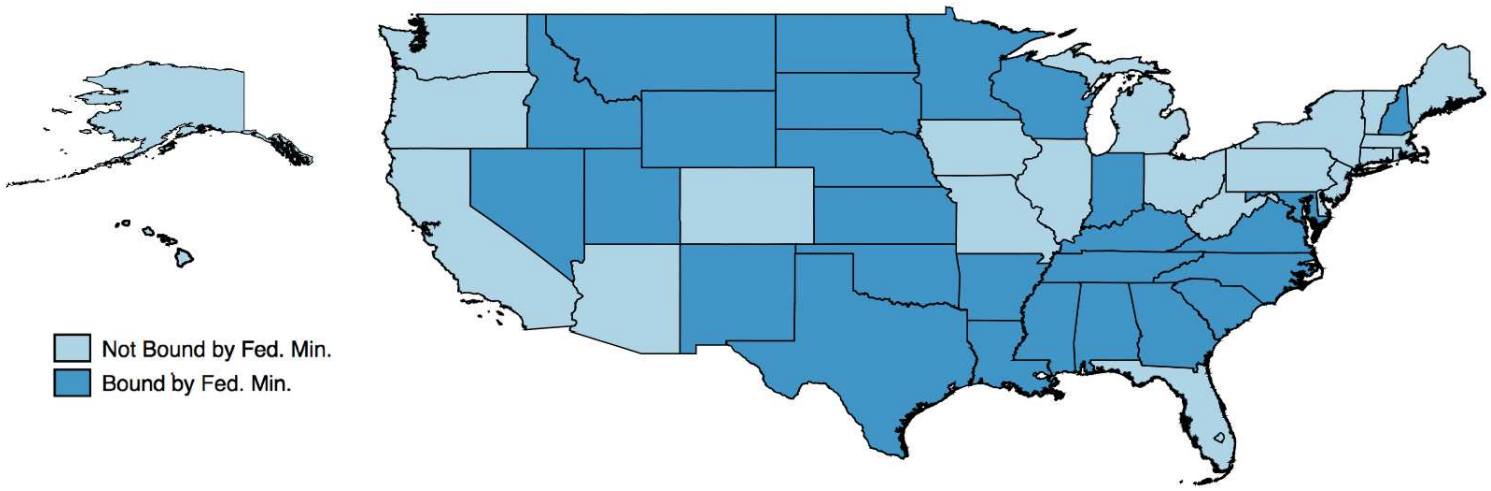


Figure 1: **States Bound by the 2008 and 2009 Federal Minimum Wage Increases:**
The map differentiates states on the basis of whether they were fully or partially bound by the July 2007, 2008 and 2009 increases in the federal minimum wage. I define states as fully bound if their January 2008 minimum wage, as reported by the Bureau of Labor Statistics (BLS), was less than $6.55. Such states were at least partially bound by the July 2008 increase in the federal minimum and fully bound by the July 2009 increase from $6.55 to $7.25.

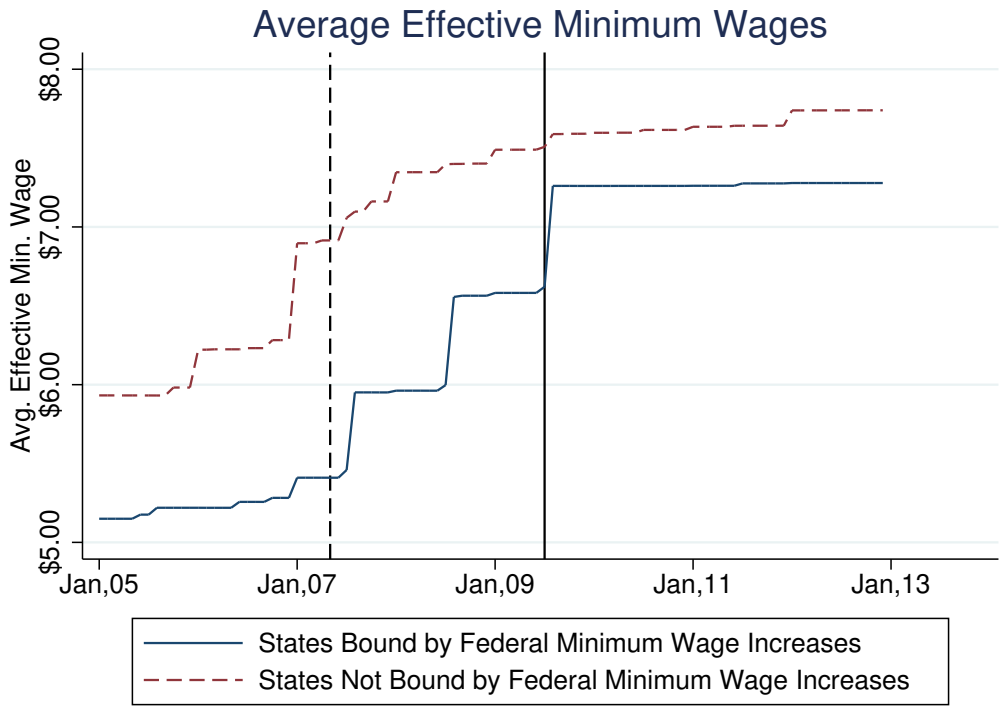


Figure 2: **Evolution of the Average Minimum Wage in Bound and Unbound States:** As in the previous figure, states are defined as fully bound if they were reported by the Bureau of Labor Statistics (BLS) to have had a minimum wage less than $6.55 in January 2008. Such states were at least partially bound by the July 2008 increase in the federal minimum and fully bound by the July 2009 increase from $6.55 to $7.25. Effective monthly minimum wage data were taken from the detailed replication materials associated with Meer and West (Forthcoming). Within each group of states, the average effective minimum wage is weighted by state population. The dashed vertical line indicates the May 2007 passage of the federal minimum wage increases, while the solid vertical line indicates the timing of the July 2009 implementation of the final increase from $6.55 to $7.25.

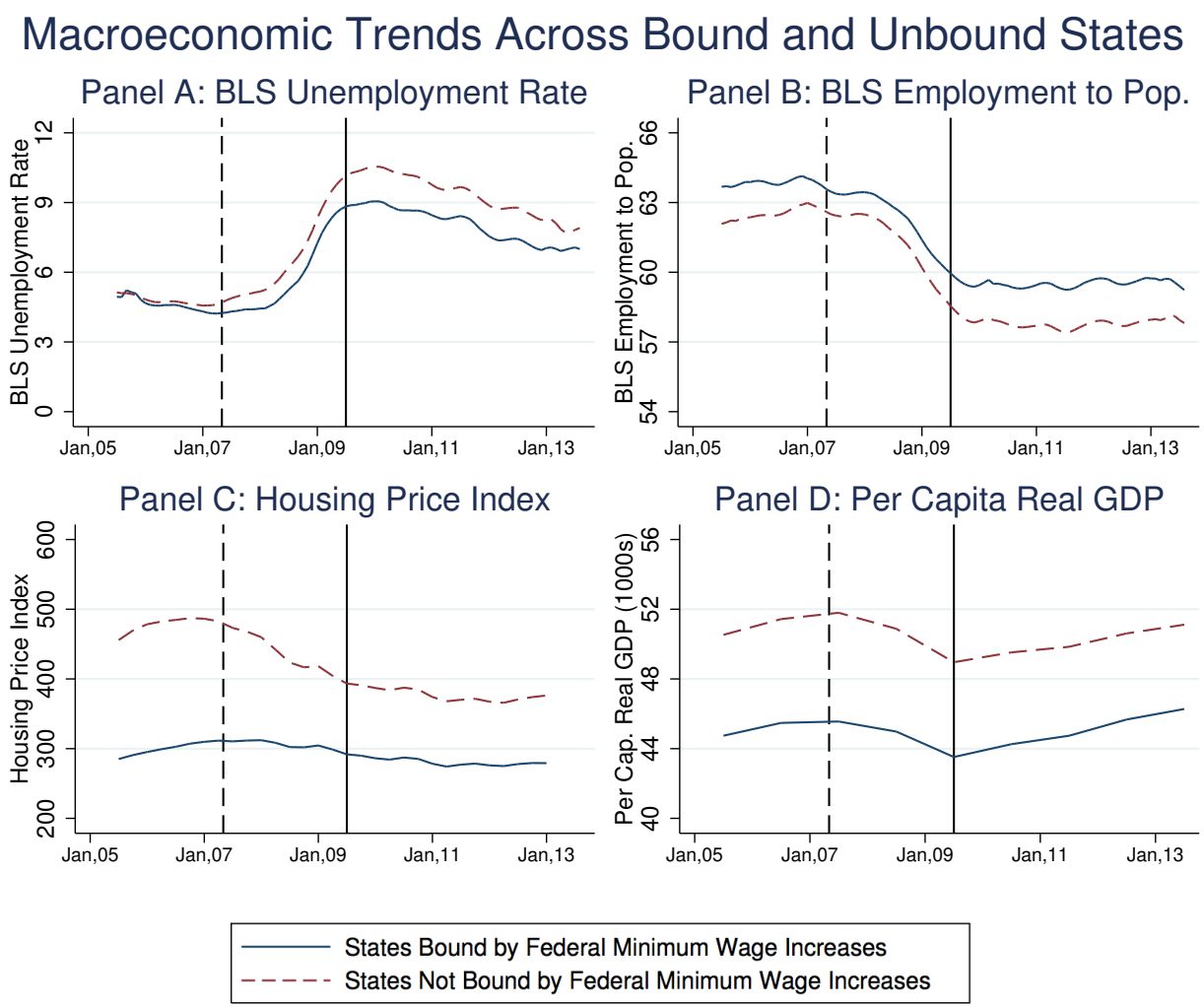


Figure 3: **Macroeconomic Trends in Bound and Unbound States:**
Bound and unbound states are defined as in previous figures. This figure's panels plot the evolution of macroeconomic indicators over the course of the housing boom and bust. All series are weighted by state population so as to reflect the weighting applied in the regression analysis. Panel A plots the average monthly unemployment rate, as reported by the BLS. Panel B plots the average monthly employment to population ratio, also as reported by the BLS. Panel C plots the average of the quarterly Federal Housing Finance Agency's all-transactions median housing price index. Panel D plots the average of annual real per capita Gross State Product, as reported by the Bureau of Economic Analysis (BEA). In each panel, the dashed vertical line indicates the May 2007 passage of the federal minimum wage increases, while the solid vertical line indicates the timing of the July 2009 implementation of the final increase from \$6.55 to \$7.25.

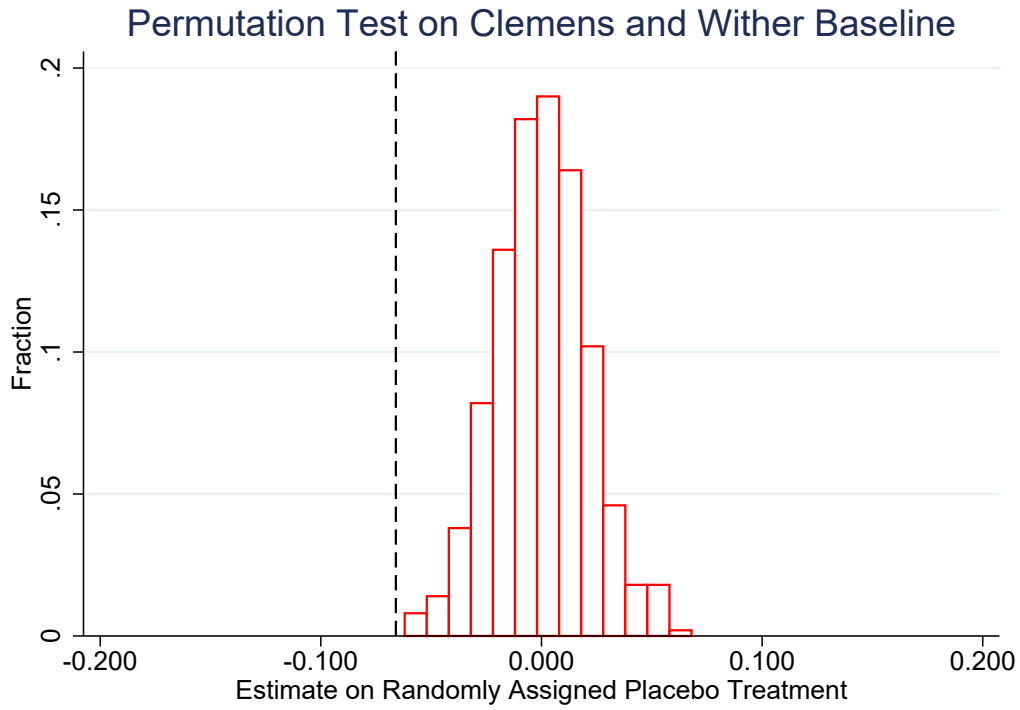
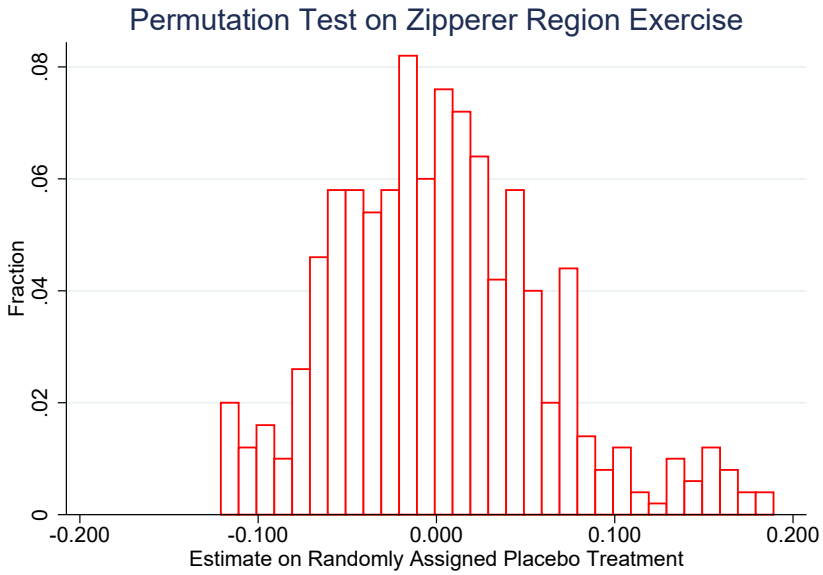


Figure 4: **Placebo Distribution for Baseline Estimate:**
Note: The figure presents the distribution of placebo point estimates associated with the baseline specification from Clemens and Wither (2014). In each iteration of the procedure, placebo treatment status was assigned to 27 of the 50 states in the sample. The baseline regression was then estimated using the placebo treatment assignment. The distribution in the figure is the distribution of estimates that resulted from 500 iterations of this procedure.

Placebo Distribution for Zipperer Falsification Test Estimates

Panel A



Panel B

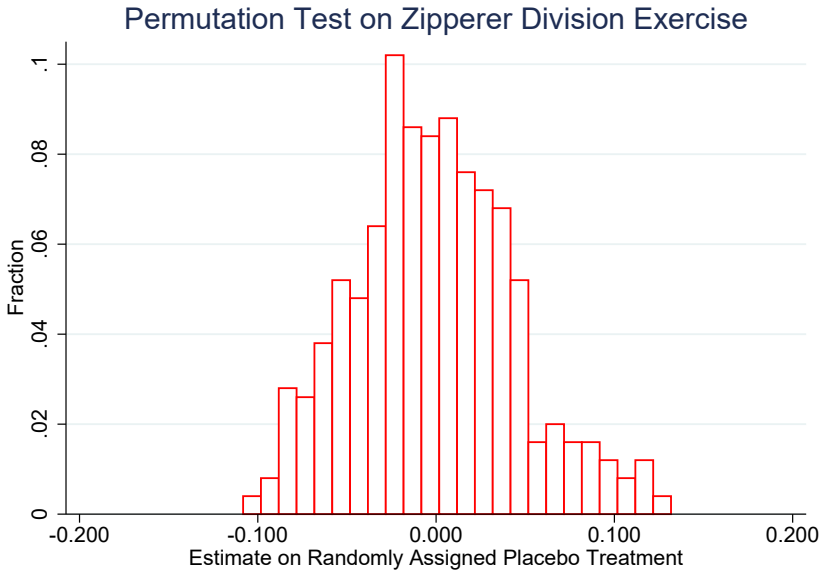


Figure 5: **Placebo Distribution for Zipperer Falsification Test Estimates:**
Note:The figure presents the distributions of placebo point estimates associated with the falsification tests from Zipperer (2016). Panel A reports the distribution associated with Zipperer's region-based exercise. In each iteration of the procedure, placebo treatment status was assigned to 3 of the 23 states in the sample Zipperer analyzed. The regression was then estimated using the placebo treatment assignment. The distribution in the figure is the distribution of estimates that resulted from 500 iterations of this procedure. Panel B reports the distribution associated with Zipperer's division-based exercise. In each iteration of the procedure, placebo treatment status was assigned to 5 of the 23 states in the sample Zipperer analyzed.

Table 1: **Reproduction of Zipperer Placebo Regression Results**

	(1)	(2)	(3)	(4)
	Zipperer Table 3		Zipperer Table A1	
	ln(minimum)	Employment	ln(minimum)	Employment
Panel A:	Census Region "Falsification Test"			
Placebo Treatment x Post 2	0.014	-0.045*	0.023+	-0.045+
	(0.011)	(0.018)	(0.012)	(0.023)
N	57,698	57,698	66,817	66,817
Panel B:	Census Division "Falsification Test"			
Placebo Treatment x Post 2	0.014	-0.056*	0.023+	-0.062*
	(0.011)	(0.024)	(0.012)	(0.027)
N	57,698	57,698	66,817	66,817
Sample	pre-2012	pre-2012	Full	Full
State and Time FE	Yes	Yes	Yes	Yes

Note: The table reproduces estimates presented in table 3 and table A.1 of Zipperer (2016).

Table 2: **Comparisons in the Severity of the Great Recession: Region-Based Exercise**

	(1)	(2)	(3)	(4)	(5)
		Average Changes		Differences	
	Plac. Treatment	Plac. Control	Rest of South	(2) - (1)	(3) - (1)
Panel A:	Comparisons of Changes from Pre-Legislation to Post-Implementation				
Income Per Cap. (1000s)	-4.03	-1.44	-0.58	2.59	3.45
Employment Rate	-0.06	-0.047	-0.041	0.013	0.019
Unemployment Rate	6.045	4.626	3.775	-1.419	-2.27
Med. House Price (Millions)	-0.161	-0.099	-0.016	0.062	0.145
Const. Output Index (BEA)	-76.69	-35.74	-22.59	40.95	54.1
Panel B:	Comparisons of Changes from Baseline to Post-Period				
Income Per Cap. (1000s)	-1.37	-0.21	-0.42	1.16	0.95
Employment Rate	-0.029	-0.027	-0.02	0.002	0.009
Unemployment Rate	1.564	1.634	1.533	0.07	-0.031
Med. House Price (Millions)	-0.06	-0.038	-0.019	0.022	0.041
Const. Output Index (BEA)	-15.39	-11.24	-9.89	4.15	5.5

Note: The table presents data that describe changes in macroeconomic conditions. In both panels, the "post" period for all changes extends from August 2009 through the end of 2012. The "pre" period from panel A consists of 2006 and January through May of 2007, which was the month during which the federal minimum wage increases were legislated. The "pre" period from panel Bconsists of August 2008 through April 2009, which corresponds with the baseline period in the SIPP regression analysis. The states described in column 1 are the District of Columbia, Florida, and West Virginia, which are the states to which Zipperer assigns "placebo treatment status" in his region-based exercise. The states described in column 2 consist of the remainder of the unbound states. The states described in column 3 consist of the states in the Southern census region that are not assigned to placebo treatment status. The macroeconomic series are further discussed in the main text. Data come from BLS, BEA, and FHFA.

Table 3: **Comparisons in the Severity of the Great Recession: Division-Based Exercise**

	(1)	(2)	(3)	(4)	(5)
		Average Changes		Differences	
	Plac. Treatment	Plac. Control	Rest of South	(2) - (1)	(3) - (1)
Panel A:	Comparisons of Changes from Pre-Legislation to Post-Implementation				
Income Per Cap. (1000s)	-3.7	-0.89	-2.02	2.81	1.68
Employment Rate	-0.059	-0.042	-0.054	0.017	0.005
Unemployment Rate	5.676	4.244	4.501	-1.432	-1.175
Med. House Price (Millions)	-0.14	-0.073	-0.046	0.067	0.094
Const. Output Index (BEA)	-76.34	-26.03	-44.48	50.31	31.86
Panel B:	Comparisons of Changes from Baseline to Post-Period				
Income Per Cap. (1000s)	-1.11	-0.20	-0.77	0.91	0.34
Employment Rate	-0.03	-0.023	-0.028	0.007	0.002
Unemployment Rate	1.692	1.577	1.572	-0.115	-0.12
Med. House Price (Millions)	-0.057	-0.029	-0.037	0.028	0.02
Const. Output Index (BEA)	-22.04	-8.53	-15.58	13.51	6.46

Note: The table presents data that describe changes in macroeconomic conditions. In both panels, the "post" period for all changes extends from August 2009 through the end of 2012. The "pre" period from panel A consists of 2006 and January through May of 2007, which was the month during which the federal minimum wage increases were legislated. The "pre" period from panel Bconsists of August 2008 through April 2009, which corresponds with the baseline period in the SIPP regression analysis. The states described in column 1 are Arizona, Colorado, the District of Columbia, Florida, and West Virginia, which are the states to which Zipperer assigns "placebo treatment status" in his division-based exercise. The states described in column 2 consist of the remainder of the unbound states. The states described in column 3 consist of the states in the South Atlantic and Mountain West census divisions that are not assigned to placebo treatment status. The macroeconomic series are further discussed in the main text. Data come from BLS, BEA, and FHFA.

Table 4: **Inference Comparisons**

	(1) Point Estimate	(2) Standard Error	(3) Width of 95 Percent CI	(4) 95 Percent CI	(5) Width of 90 Percent CI	(6) 90 Percent CI
Panel A:			Clemens and Wither Baseline			
Clustered Robust SE	-0.066	0.02	0.078	[-0.039, 0.039]	0.066	[-0.033, 0.033]
Permutation Test	-0.066	n/a	0.087	[-0.044, 0.043]	0.069	[-0.036, 0.033]
Panel B:			Zipperer's Region-Based "Falsification Test"			
Clustered Robust SE	-0.045	0.018	0.071	[-0.035, 0.035]	0.059	[-0.030, 0.030]
Permutation Test	-0.045	n/a	0.255	[-0.106, 0.149]	0.196	[-0.087, 0.109]
Panel C:			Zipperer's Division-Based "Falsification Test"			
Clustered Robust SE	-0.056	0.024	0.094	[-0.047, 0.047]	0.079	[-0.040, 0.040]
Permutation Test	-0.056	n/a	0.180	[-0.082, 0.098]	0.156	[-0.072, 0.084]

Note: The table reports confidence intervals estimated using both clustered robust standard errors and permutation test methods. The confidence intervals reported in panel A are those that apply to the baseline estimates from Clemens and Wither (2014). The estimates in panel B are those that apply to Zipperer's region-based falsification test. The estimates in panel C are those that apply to Zipperer's division-based falsification test. The estimates reveal that clustered robust standard errors yield confidence intervals of essentially the same size as placebo test methods for the Clemens and Wither baseline. By contrast, clustered robust standard errors understate the confidence interval associated with Zipperer's region-based test by a factor of more than 3. Clustered robust standard errors understate the confidence interval associated with Zipperer's division-based test by a factor of just under 2.

Table 5: **Sensitivity of Zipperer Falsification Test Estimates**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dependent Variable: Employed (Reduced Form)							
Panel A:								
Placebo Bound x Post 1	-0.062	0.108**	0.102**	-0.070*	-0.061+	0.046	0.042	-0.058+
	(0.037)	(0.028)	(0.028)	(0.026)	(0.034)	(0.041)	(0.041)	(0.029)
Placebo Bound x Post 2	-0.044*	0.020	0.021	-0.026	-0.056*	-0.053	-0.050	-0.043
	(0.018)	(0.042)	(0.043)	(0.031)	(0.024)	(0.048)	(0.048)	(0.041)
Weighted	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel B:								
Placebo Bound x Post 1	-0.060	0.091***	0.083**	-0.060+	-0.055	0.038	0.033	-0.055+
	(0.040)	(0.022)	(0.023)	(0.030)	(0.036)	(0.043)	(0.043)	(0.030)
Placebo Bound x Post 2	-0.054**	-0.026	-0.028	-0.017	-0.062**	-0.071+	-0.069	-0.048
	(0.015)	(0.033)	(0.035)	(0.031)	(0.021)	(0.041)	(0.041)	(0.041)
Weighted	No	No	No	No	No	No	No	No
N	57,702	48,814	35,339	57,702	57,702	48,814	35,339	57,702
Placebo Determination	Region	Region	Region	Region	Division	Division	Division	Division
Match Criterion	None	No FL-AZ	20K	None	None	No FL-AZ	20K	None
More Macro Controls	No	No	No	Yes	No	No	No	Yes

Note: +, *, **, and *** indicate statistical significance at the 0.10, 0.05, 0.01, and 0.001 levels respectively. Columns 1 and 5 present estimates of equation (3), which is Zipperer's falsification test, on the full sample of unbound states. Columns 2 and 5 report estimates of equation (3) on a sample that excludes Florida and Arizona, which experienced extreme housing declines, from the set of unbound states. Columns 3 and 6 report estimates of equation (3) on a sample that all states that were difficult to matched on the basis of their housing declines. Columns 4 and 8 report estimates in which equation (3) is augmented to include additional controls for variations in macroeconomic conditions. The controls include the aggregate employment rate, unemployment rate, personal income per capita, and the housing price index, each of which are interacted with a set of period fixed effects. Standard errors allow for correlation clusters across errors at the state level.

Table 6: **Sensitivity of Zipperer Falsification Test Estimates**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dependent Variable: Employed (Reduced Form)							
Panel A:								
Placebo Bound x Post 1	-0.054	0.152***	0.149***	-0.049+	-0.066+	0.062	0.059	-0.066*
	(0.042)	(0.029)	(0.031)	(0.027)	(0.035)	(0.059)	(0.060)	(0.028)
Placebo Bound x Post 2	-0.027	0.088	0.085	0.004	-0.052	-0.032	-0.034	-0.045
	(0.025)	(0.062)	(0.064)	(0.034)	(0.030)	(0.075)	(0.076)	(0.055)
Weighted	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel B:								
Placebo Bound x Post 1	-0.066	0.109***	0.103**	-0.047	-0.075+	0.041	0.037	-0.075*
	(0.044)	(0.025)	(0.029)	(0.035)	(0.036)	(0.054)	(0.054)	(0.030)
Placebo Bound x Post 2	-0.050*	0.026	0.018	0.002	-0.069*	-0.055	-0.060	-0.058
	(0.023)	(0.048)	(0.050)	(0.035)	(0.028)	(0.062)	(0.062)	(0.056)
Weighted	No	No	No	No	No	No	No	No
N	57,583	48,736	35,257	57,583	57,583	48,736	35,257	57,583
Placebo Determination	Region	Region	Region	Region	Division	Division	Division	Division
Match Criterion	None	No FL-AZ	20K	None	None	No FL-AZ	20K	None
More Macro Controls	No	No	No	Yes	No	No	No	Yes

Note: +, *, **, and *** indicate statistical significance at the 0.10, 0.05, 0.01, and 0.001 levels respectively. Columns 1 and 5 present estimates of equation (3), which is Zipperer's falsification test, on the full sample of unbound states. Columns 2 and 5 report estimates of equation (3) on a sample that excludes Florida and Arizona, which experienced extreme housing declines, from the set of unbound states. Columns 3 and 6 report estimates of equation (3) on a sample that all states that were difficult to matched on the basis of their housing declines. Columns 4 and 8 report estimates in which equation (3) is augmented to include additional controls for variations in macroeconomic conditions. The controls include the aggregate employment rate, unemployment rate, personal income per capita, and the housing price index, each of which are interacted with a set of period fixed effects. Standard errors allow for correlation clusters across errors at the state level.

Table 7: **Does Zipperer's Preferred Specification Pass His Own Falsification Test?**

	(1)	(2)	(3)	(4)
Panel A:	Baseline	Zipperer	Zipperer Falsification	
Bound x Post 1	-0.039*	-0.028		
	(0.018)	(0.019)		
Bound x Post 2	-0.059**	-0.024		
	(0.021)	(0.024)		
Placebo Bound x Post 1			0.032	0.144*
			(0.045)	(0.053)
Placebo Bound x Post 2			0.068	0.148+
			(0.058)	(0.076)
Weighted	Yes	Yes	Yes	Yes
Panel B:	Baseline	Zipperer	Zipperer Falsification	
Bound x Post 1	-0.044*	-0.041*		
	(0.019)	(0.019)		
Bound x Post 2	-0.066**	-0.043*		
	(0.020)	(0.021)		
Placebo Bound x Post 1			0.026	0.117+
			(0.048)	(0.063)
Placebo Bound x Post 2			0.059	0.126
			(0.062)	(0.079)
Weighted	No	No	No	No
N	147,459	147,459	57,583	57,583
Sample of States	Unbound	Unbound	Unbound	Unbound
Placebo Determination	NA	NA	Region	Division
Ind. Share x Time FE	No	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes

Note: +, *, **, and *** indicate statistical significance at the 0.10, 0.05, 0.01, and 0.001 levels respectively. Column 1 presents an estimate of equation (1). Column 2 presents an estimate of equation (2). Columns 3 and 4 present falsification tests of the same form as equation (3), but for which the underlying regression is that from equation (2). The estimates thus show that applying Zipperer's falsification test to his preferred specification yields economically large and positive values. Under Zipperer's interpretation of the falsification test, this implies significant bias against estimating negative employment effects associated with the minimum wage changes under analysis. Standard errors allow for correlation clusters across errors at the state level.