ORIGINAL ARTICLE   **Open Access**

More on recent evidence on the effects of minimum wages in the United States

David Neumark[1]*, JM Ian Salas[2] and William Wascher[3]

* Correspondence: dneumark@uci.edu
[1]UCI, NBER, and IZA, 3151 Social Science Plaza, Irvine, CA 92697, USA
Full list of author information is available at the end of the article

Abstract

A central issue in estimating the employment effects of minimum wages is the appropriate comparison group for states (or other regions) that adopt or increase the minimum wage. In recent research, Dube et al. (Rev Econ Stat 92:945-964, 2010) and Allegretto et al. (Ind Relat 50:205-240, 2011) argue that past U.S. research is flawed because it does not restrict comparison areas to those that are geographically proximate and fails to control for changes in low-skill labor markets that are correlated with minimum wage increases. They argue that using "local controls" establishes that higher minimum wages do not reduce employment of less-skilled workers. In Neumark et al. (Ind Labor Relat Rev 67:608-648, 2014), we present evidence that their methods fail to isolate more reliable identifying information and lead to incorrect conclusions. Moreover, for subsets of treatment groups where the identifying variation they use is supported by the data, the evidence is consistent with past findings of disemployment effects. Allegretto SA, Dube A, Reich M, Zipperer B (2013a) Credible research designs for minimum wage studies. IZA Discussion Paper No. 7638, Bonn, Germany have challenged our conclusions, continuing the debate regarding some key issues regarding choosing comparison groups for estimating minimum wage effects. We explain these issues and evaluate the evidence. In general, we find little basis for their analyses and conclusions and argue that the best evidence still points to job loss from minimum wages for very low-skilled workers – in particular, for teens.

JEL codes: J23; J38
Keywords: Minimum wages; Employment; Comparison groups

Introduction

Recent debate on the employment effects of minimum wages has focused on the proper specification of the control groups for estimating the effects of minimum wages. This is a long-standing issue that has been confronted in different ways beginning with research in the early part of the last century (Neumark et al., 2014). In the current incarnation of this debate, Dube et al. (2010, hereafter DLR) and Allegretto et al. (2011, hereafter ADR) have argued that to obtain valid estimates of minimum wage effects it is essential to control for "spatial heterogeneity", using nearby geographic areas as controls to better account for where and when minimum wages are adopted or increased. Moreover, based on their specifications of how best to control for this spatial heterogeneity, DLR and ADR have put forward a severe critique of the findings from much of the existing U.S. evidence on the employment effects of minimum wages using state-level panel data. They argue that this evidence is biased because of "a spurious negative relationship between the minimum wage and employment for low wage

workers…" (Dube 2011, p. 763), owing to minimum wages being adopted when there are negative shocks to the employment of affected workers unrelated to minimum wage effects.

Their evidence in support of this claim comes from implementing two types of what they refer to as "local controls". The first is the inclusion in their regression models of jurisdiction-specific linear time trends. The second is the inclusion of interactions between period dummy variables and dummy variables for sets of nearby states or neighboring counties, so that minimum wage effects are identified only net of common changes within these sets. Based on these two approaches, they argue (in DLR, for example) that there are "no detectable employment losses from the kind of minimum wage increases we have seen in the United States" (p. 962).

In Neumark et al. (2014, hereafter NSW), we presented evidence that the methods advocated in these studies do not isolate more reliable identifying information (i.e., better comparison groups) and thus are flawed and lead to incorrect conclusions. In one case – the issue of state-specific trends – we explicitly demonstrate the problem with their methods and show how more appropriate ways of controlling for unobserved trends that affect teen employment lead to evidence of disemployment effects similar to past studies. In the other case – identifying minimum wage effects from the variation within Census divisions or, even more narrowly, within contiguous cross-border county pairs – we show that the exclusion of other regions or counties as potential controls is generally not supported by the data. Moreover, for regions where restricting the identifying variation in this way *is* supported by the data, the evidence is consistent with past findings of disemployment effects. Finally, when we let the data determine the appropriate control states to use for estimating the effects of state minimum wage increases in the Current Population Survey (CPS) data, we find evidence of disemployment effects for teens, with elasticities near –0.15.

Most recently, Allegretto et al. (2013a, hereafter ADRZ) have criticized our conclusions, attacking much of our evidence. In the present paper, we lay out several issues that we see as forming the crux of this debate about the use of local controls to construct better comparison groups and provide our own analysis of these issues. In general, we find little basis for the alternative analyses and conclusions ADRZ present, and we conclude that the best evidence still points to job loss from minimum wages for low-skilled workers – in particular for teens.[1]

In our view, the key issue is whether the identifying assumptions entailed by ADR's and DLR's use of local controls lead to more biased or less biased estimates of the employment effect of minimum wages. Our previous work concluded that the assumptions implicit in their methods bias the estimated effect toward zero. The issue was *not* simply whether the variation left unused by DLR and ADR reduced the efficiency of their estimates of the employment effects of minimum wages. Labor economists often use approaches thought to reduce bias at the cost of less precise estimates. In the final section of this paper we provide a possible explanation for why limiting attention to local controls might produce a bias towards finding no employment effect and cite recent evidence from Baskaya and Rubinstein (2012) consistent with this explanation. At the same time, estimation of richer specifications prompted by some of the analyses presented by ADRZ undermines the conclusion that, for teens, including local controls reduces the estimated employment effect of the minimum wage. So in either case, the claim that controlling for spatial heterogeneity establishes that minimum wages do not reduce employment of unskilled workers is unfounded.[2]

Issue 1: Using synthetic control methods to construct local controls

The central element of the papers by ADR and DLR is their use of specifications that identify effects only from states in the same Census division (in ADR) or from pairs of contiguous counties straddling state borders (in DLR). The implicit assumption in these specifications is that geographically proximate areas provide better controls. However, as discussed in NSW, one can actually test this assumption using tools borrowed from the synthetic control approach to estimating treatment effects (Abadie et al., 2010). And in the context of the ADR and DLR studies, we showed that the weight put on nearby states or counties as potential controls (or "donors", in the language of the synthetic control literature) for states or counties in which minimum wages increased were generally no higher than the weight put on states or counties farther away, and indeed that these nearby states or counties tended to get no more weight than a randomly-selected state or county.

In their response, ADRZ claim that we have glossed over an important conceptual issue – namely, that "examining weights within Census divisions and comparing these to weights outside divisions does not tell us whether comparing local areas is better than using state panel regressions with two-way fixed effects" (p. 63). However, we did not present the results of our synthetic control analysis as an explicit validation of the standard two-way fixed effects estimator, and by setting up this straw man, ADRZ distract attention from our main point: The synthetic control analysis *is* informative about whether to focus on local controls or not because it tells us whether it makes sense to put all the weight on the within-region variation (Census divisions in ADR and cross-border county pairs in DLR) or instead to put weight on variation from outside the region as well (however that weight might be distributed). In particular, for the analysis of state-level data, we reported that the average weight per same-division donor state is higher than 1/(number of potential donors) in only 18 of 50 cases; that is, in more than 60 percent of cases, the average weight on same-division donor states based on the synthetic control matching is less than the weight we would get if all potential donors were weighted equally.

Weights on same-division states vs. other states

ADRZ also dispute our calculation of the weights, arguing (focusing on the CPS analysis) that the evidence actually shows that, "A donor state within the same division receives weights that are *2.8 to 4.1 times as large* as weights for donors outside of the division" (p. 66). For example, for our matching on regression residuals in Table 3 of NSW, they calculate that the average weight per donor state in the same Census division is 0.098, versus 0.035 for other states, for a ratio of 2.806 (the source for their first number cited above; see their Table B1). Thus, ADRZ conclude, "a straightforward interpretation of NSW's own evidence indicates that neighboring areas are more alike than are places farther away – contradicting their central thesis" (p. 66).

This is a direct contradiction of the results we report – i.e., that the weight on same-division states is generally no higher than the weight on states in other divisions. But ADRZ's conclusion is based on a flawed calculation that weights states in a manner that mechanically tends to produce a high ratio of the weight they compute on same-division versus non-same-division states.[3]

To see this, let p_{ij}^{S} denote the weight put on state i in treatment j for the same-division states, let N_j^{S} denote the number of same-division states in treatment j, and let T denote the number of treatments.[4] ADRZ's calculation for same-division states is

$$
\frac{\sum_j\left(\sum_i p_{ij}^{s}\right)/T}{\sum_j\left(\sum_i N_j^{s}\right)/T} \; = \; \frac{\sum_j\left(\sum_i p_{ij}^{s}\right)}{\sum_j\left(\sum_i N_j^{s}\right)} \tag{1}
$$

That is, they add up all the weight on same-division states across all the treatments, and divide that by the number of treatments, and divide that by the total number of same-division states in all the treatments, also dividing that by the number of treatments. They then do the same calculation for other-division states and compute the ratio of the two.

This calculation puts very high weight on the treatments with large number of donors. In the data, the number of donors varies widely across treatments, and the number of other-division donors can be very large. The number of same-division donors ranges from 1 to 8, with a standard deviation of 2, while the number of non-same-division donors varies from 1 to 45, with a standard deviation of 18. As a result, ADRZ's calculation is particularly sensitive to the observations on other-division donors from treatments with large numbers of such donors. Because the number of other-division donors can be so much higher, the ratio of the expression in equation (1) for same-division relative to other-division states tends to get blown up by this feature of ADRZ's calculation.

The top panel of Table 1 provides an illustrative example. The table shows the number of donors in the same and the other divisions in each of five hypothetical treatments and the weights each state gets in the hypothetical synthetic control analysis. In this example, there are four treatments with the same number of same-division and other-division donors (two of each). In these four treatments, the weight on each same-division state (0.24) is slightly less than the weight on each other-division state (0.26). In the fifth treatment, there are also two donors in the same-division states, each with weight of 0.02, and a large number of non-donor states (48) – mimicking what actually happens in the data – each with the same weight as the same-division states (0.02).

If these weights resulted from a synthetic control analysis like the one we proposed, what would we conclude? In four of the five treatments, the average weight on other-division states is higher (0.26 vs. 0.24), while in the fifth the weight on same-division states is the same (0.02). We would argue that the interpretation of this kind of evidence from a synthetic control analysis would be similar to the interpretation in NSW: There is no strong evidence that more weight – let alone *all* the weight – should go on same-division states.

But what does ADRZ's calculation suggest? Using equation (1) and its equivalent version for other-division states, the resulting value is 3.61 (reported in the table), within the range they use to conclude that same-division states are much more alike – i.e., better controls – than other-division states. Yet looking at the weights for example 1 in Table 1, this does not seem a supportable conclusion.

One could use the unweighted average of the ratio of the weights on same-division to other-division states, which would be equal to 0.938, indicating slightly lower weight on same-division states, which seems like the right answer. Alternatively, if one wanted to use a calculation more comparable to the calculation ADRZ claim to present –

Table 1 Examples of weights on same-division states and other-division states

	Same division		Other division	
	Weight per state	# states	Weight per state	# states
Example 1				
Treatment 1	.24	2	.26	2
Treatment 2	.24	2	.26	2
Treatment 3	.24	2	.26	2
Treatment 4	.24	2	.26	2
Treatment 5	.02	2	.02	48
ADRZ calculation (equation (1))	$\dfrac{\left\{\frac{8\times0.24+2\times0.02}{5}\right\}/\left\{\frac{10}{5}\right\}}{\left\{\frac{8\times0.26+48\times0.02}{5}\right\}/\left\{\frac{56}{5}\right\}}=3.61$			
Equation (2) calculation	$\dfrac{\left\{4\times\left(\frac{0.48}{2}\right)+1\times\left(\frac{0.04}{2}\right)\right\}/5}{\left\{4\times\left(\frac{0.52}{2}\right)+1\times\left(\frac{0.96}{48}\right)\right\}/5}=0.925$			
Example 2				
Treatment 1	.24	2	.26	2
Treatment 2	.24	2	.26	2
Treatment 3	.24	2	.26	2
Treatment 4	.24	2	.26	2
Treatment 5	.01	4	.01	96
ADRZ calculation (equation (1))	$\dfrac{\left\{\frac{8\times0.24+4\times0.01}{5}\right\}/\left\{\frac{12}{5}\right\}}{\left\{\frac{8\times0.26+96\times0.01}{5}\right\}/\left\{\frac{104}{5}\right\}}=5.59$			
Equation (2) calculation	$\dfrac{\left\{4\times\left(\frac{0.48}{2}\right)+1\times\left(\frac{0.04}{4}\right)\right\}/5}{\left\{4\times\left(\frac{0.52}{2}\right)+1\times\left(\frac{0.96}{96}\right)\right\}/5}=0.924$			

"average per-donor weight of same-division donors, relative to the per-donor weight of the other-division donors" (p. 65) – one would want to use the following equation for same-division controls:

$$\sum_j \left[\left(\sum_i p_{ij}^S\right)/N_j^S\right]/T \tag{2}$$

and the corresponding equation for other-division controls. In that case, example 1 yields the number 0.925 (reported in the table). Thus, the number resulting from ADRZ's calculation seems much too high.[5]

Example 2 in Table 1 makes it even clearer that ADRZ's calculation is flawed. In this example, we simply modify treatment 5 so that there are twice as many same-division donors and twice as many other-division donors – and correspondingly we cut the weight on each state in half. It seems obvious to us that the conclusion one draws about appropriate donors from example 2 should be the same as example 1: There are still four of five treatments for which the weight on other-division states is higher. Yet because the ADRZ calculation upweights the large number of donors, the resulting number increases by more than 50 percent, from 3.61 to 5.59. In contrast, the calculation using equation (2) scarcely changes. Finally, *both* of the numbers resulting from ADRZ's calculation far exceed one, even though the weight on same-division states is equal to or less than the weight on other-division states for every treatment.

Thus, our examples and this discussion highlight two conclusions: (1) that the number resulting from ADRZ's calculation does not have a sensible interpretation with regard to which states are the appropriate controls; and (2) that it is highly sensitive to differences that have no bearing on the evaluation of same-division and other-division states as controls. We therefore stand by the conclusion from our synthetic control analysis that there is little or no evidence to indicate that same-division states are better controls than other-division states, and certainly no evidence that the latter should be excluded as controls.

ADRZ also present a synthetic control analysis that does not use minimum wage increases to identify treatment observations, but randomly assigns a placebo minimum wage law to an individual state in a time period and then calculates the synthetic control donor weights for all remaining states. They suggest that this approach is informative because it "dispenses with the shortcomings" (p. 34) of the kind of analysis we did, by which they mean that we could use only a subset of minimum wage increases as treatments for the synthetic control analysis. Their Figure eight shows that their computed weights decline monotonically with distance from the treated state to the donor state, up to about 1,000 miles (and then are flat). This evidence, they argue, "unambiguously demonstrates that the synthetic control algorithm assigns much greater weight to nearby states when constructing the counterfactual teen employment" (p. 34).

However, this approach strikes us as uninformative about the question at hand – whether a particular subset of states provides a more valid set of controls for states where the minimum wage actually does increase. Since the whole point of the approach taken in ADR and DLR is their presumption that actual minimum wage increases are associated with the residuals of the estimated employment regressions (either because of policy endogeneity or coincidence), we want to know precisely whether the nearby states provide better controls for these treatment observations. As we have already discussed and in contrast to ADRZ's claims, the data indicate that for the minimum wage increases observed in the data, the same-division states do not provide better controls. We present additional evidence on this below.

Treatments used in the synthetic control analysis

We also present some analyses that attempt to use the synthetic control estimator to identify control observations and then estimate the effects of minimum wages on employment based on those controls. ADRZ criticize our matching estimator because the subset of "clean" minimum wage increases – treatments with donors that have no minimum wage increases in the previous four quarters and the following three – does not produce a negative and significant minimum wage effect (the estimated elasticity is only around –0.06 in the state-level data). In particular, since we suggested in our paper that the fact that we do not replicate the standard panel data estimator using this subset of minimum wage increases made this subset of minimum wage increases "unusual," they question whether it is valid to use this subset of increases to assess the plausibility of restricting attention to neighboring states as controls, as we do in our synthetic control analysis. We think it is useful to assess whether ADR and DLR throw out states or counties that are valid controls for the subset of minimum wage increases for which the analysis is cleanest. However, the answer does not depend on restricting attention to these minimum wage increases.

Table 2 Mean synthetic control weights per state in same division and other divisions, CPS data at state by quarter level, 1990 – 2011:Q2

	Mean synthetic control weight per same-division state					Mean synthetic control weight per other-division state				
	Residual	*Residual, 0 MW effect*	*Teen empl. level*	*Teen empl., 1-quarter diff.*	*Teen empl., 4-quarter diff.*	*Residual*	*Residual, 0 MW effect*	*Teen empl. level*	*Teen empl., 1-quarter diff.*	*Teen empl., 4-quarter diff.*
Matched on:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
New England	0.029	0.025	0.038	0.032	0.016	0.019	0.019	0.018	0.019	0.020
Mid-Atlantic	0.015	0.013	0.026	0.015	0.010	0.020	0.020	0.020	0.020	0.020
East North Central	0.010	0.010	0.008	0.010	0.016	0.021	0.021	0.021	0.021	0.020
West North Central	0.021	0.024	0.098	0.022	0.015	0.020	0.019	0.009	0.020	0.021
South Atlantic	0.027	0.027	0.022	0.020	0.026	0.019	0.019	0.020	0.020	0.019
East South Central	0.039	0.040	0.029	0.041	0.025	0.019	0.019	0.019	0.019	0.020
West South Central	0.011	0.010	0.024	0.015	0.015	0.021	0.021	0.020	0.020	0.020
Mountain	0.017	0.017	0.023	0.027	0.025	0.020	0.021	0.020	0.019	0.019
Pacific	0.024	0.025	0.031	0.020	0.021	0.020	0.020	0.019	0.020	0.020
All	0.022	0.022	0.035	0.023	0.020	0.020	0.020	0.018	0.020	0.020

Notes: Table reports calculations from synthetic control matching on all minimum wage increases in the sample period. The donor pool used consists of all states other than the treatment state. The estimates in columns (1), (2), (6), and (7) are based on residuals from panel data estimates of the log of the teen employment rate on fixed state and period effects and controls for the aggregate unemployment rate and the relative size of the teen population. The log of the minimum wage is included in the regression used in columns (1) and (6). (See NSW, 2014, for more details).

Table 3 Mean synthetic control weights per county for contiguous and non-contiguous counties, county-level QCEW data, 1990–2006:Q2

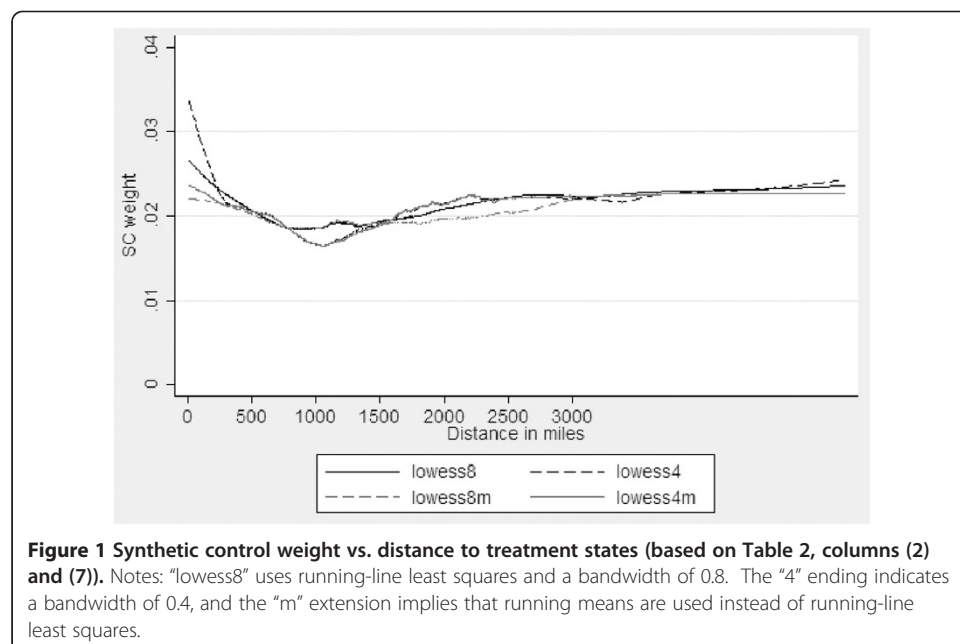| Matched on:	Mean synthetic control weight per contiguous county					Mean synthetic control weight per non-contiguous county				
	Residual	Residual, 0 MW effect	Rest. empl. level	Rest. empl., 1-quarter diff.	Rest. empl., 4-quarter diff.	Residual	Residual, 0 MW effect	Rest. empl. level	Rest. empl., 1-quarter diff.	Rest. empl., 4-quarter diff.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Minimum	0.000	0.000	0.000	0.000	0.000	0.005	0.008	0.007	0.007	0.013
10th percentile	0.000	0.000	0.000	0.000	0.000	0.019	0.019	0.019	0.019	0.019
25th percentile	0.002	0.002	0.001	0.002	0.002	0.020	0.020	0.020	0.020	0.020
Median	0.006	0.006	0.006	0.005	0.005	0.020	0.020	0.020	0.020	0.020
75th percentile	0.011	0.011	0.012	0.010	0.010	0.020	0.020	0.020	0.020	0.020
90th percentile	0.024	0.024	0.026	0.023	0.019	0.020	0.020	0.020	0.020	0.020
Maximum	0.484	0.493	0.336	0.412	0.219	0.021	0.021	0.021	0.021	0.020
Mean	0.011	0.011	0.012	0.011	0.009	0.020	0.020	0.020	0.020	0.020

Notes: Table reports calculations from synthetic control matching on all minimum wage increases in the sample period. The donor pool used consists of the top 50 counties with the lowest RMSPE from all states other than the treatment state, adding in contiguous cross-border counties if they are not already in this set of 50. The estimates in columns (1), (2), (6), and (7) are based on residuals from panel data estimates of the log of the ratio of restaurant employment to county population on fixed county and period effects and controls for the log of the ratio of aggregate private-sector employment to county population. The log of the minimum wage is included in the regression used in columns (1) and (6). (See NSW, 2014, for more details).
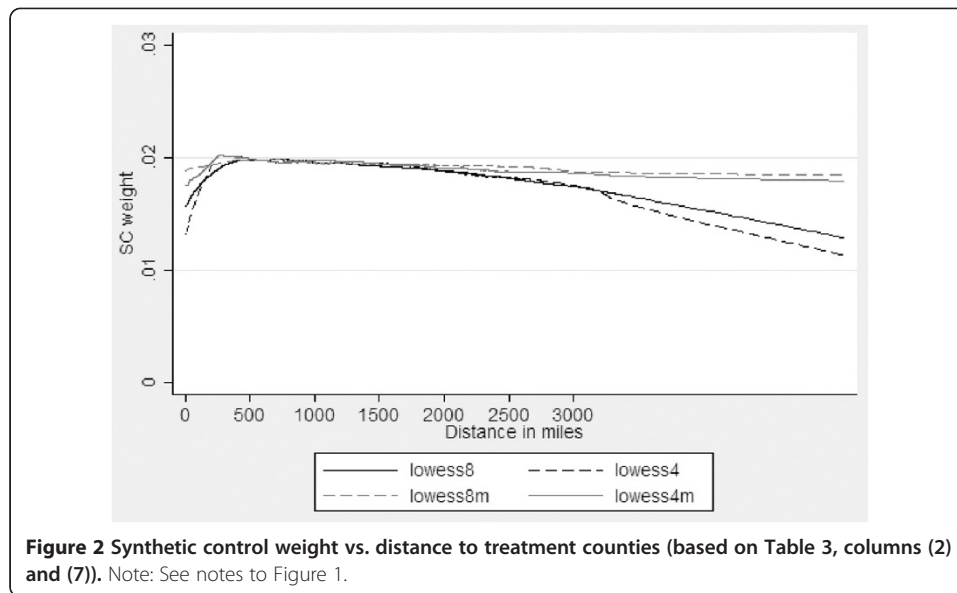
In Tables 2 and 3, we show estimates from the synthetic control exercise used in Tables 3 and 5 of NSW, with the matching now done on all observations where there was a minimum wage change. Clearly this is more problematic because the frequency of minimum wage changes implies that "donors" can be contaminated in either the pre- or post-treatment period relative to any treatment state. For that reason, we think the most informative matching is on residuals from the standard employment equation, to account for this minimum wage variation – although this raises the issue of what estimated effect of the minimum wage to use in computing these residuals. We therefore follow what we did in NSW and report these estimates matching on residuals based on the standard panel data estimates, as well as estimates in which we zero out the effect of the minimum wage. These two alternatives can be interpreted as covering the range of most estimates in the debate to date.[6]

Table 2 reports the results for teen employment using the CPS data. Comparisons of columns (1)-(5) with columns (6)-(10) show that almost without exception there is very little basis for restricting attention to the same-division states as controls. The per-state synthetic control weights – computed appropriately, as discussed earlier – are generally quite similar for the same-division and other-division states.[7] Table 3 reports the county-level analysis. Here, there is even more compelling evidence that contiguous counties are not better controls, as the per-county synthetic control weight is generally larger for non-contiguous counties.

Figures 1 and 2 report additional information on how the weight assigned to "control" states (or counties) from this analysis varies with distance from the minimum wage increase in question – similar to what ADRZ did in their paper, but now focusing on actual minimum wage variation rather than a randomly-assigned placebo.[8] For the analysis that matches on residuals from the specification that restricts the minimum wage effect to be zero, Figure 1 shows that there is a modest increase in the synthetic control weight as one gets closer to a treated state – although a much smaller ascent


Figure 1 Synthetic control weight vs. distance to treatment states (based on Table 2, columns (2) and (7)). Notes: "lowess8" uses running-line least squares and a bandwidth of 0.8. The "4" ending indicates a bandwidth of 0.4, and the "m" extension implies that running means are used instead of running-line least squares.

Figure 2 Synthetic control weight vs. distance to treatment counties (based on Table 3, columns (2) and (7)). Note: See notes to Figure 1.

than what ADRZ suggested.[9] And Figure 2 reports the similar graph for counties, where the weight is actually *lower* for the closer counties. (ADRZ did not report any results along these lines for counties).

Issue 2: Pre-trends as evidence of spatial heterogeneity

ADRZ assert that the standard panel data model with only fixed state and year effects exhibits spurious pre-trends, with large negative leading effects of minimum wages up to three years prior to minimum wage increases (their Figures 6 and B1). Moreover, they argue that the models with division × quarter fixed effects and state-specific linear trends do not exhibit these pre-trends. They use this claim to bolster their general contention about spatial heterogeneity (although, of course, more relevant is whether there are negative shocks to youth labor markets *contemporaneous with* minimum wage increases).

However, we have been unable to replicate these results from ADRZ; indeed, we get quite different answers.[10] As a preliminary, we would make two points. First, given that our earlier paper raised sufficient doubts about the inclusion of linear state trends, we believe the key question is whether the addition of the division × quarter interactions are needed. Moreover, we would argue that the region × period interactions represent the central innovation that ADR and DLR proposed and provide an easily interpretable alternative (and restricted) identification strategy based solely on within-region variation. Hence, we focus on the role of these spatial controls.[11] Second, ADRZ provide figures that appear to be based on models with leads only. Although it does not bear on our inability to replicate their results, we would argue that it is more appropriate to estimate models that include the possible real effects of minimum wages – i.e., those that occur contemporaneously or with a lag – if for no other reason than that the actual contemporaneous or lagged effects would otherwise be omitted variables.[12]

Panel A of Figure 3 reports results where we include leads and lags of up to three years. The estimates shown are the cumulative sum of these, beginning at a lead of 12

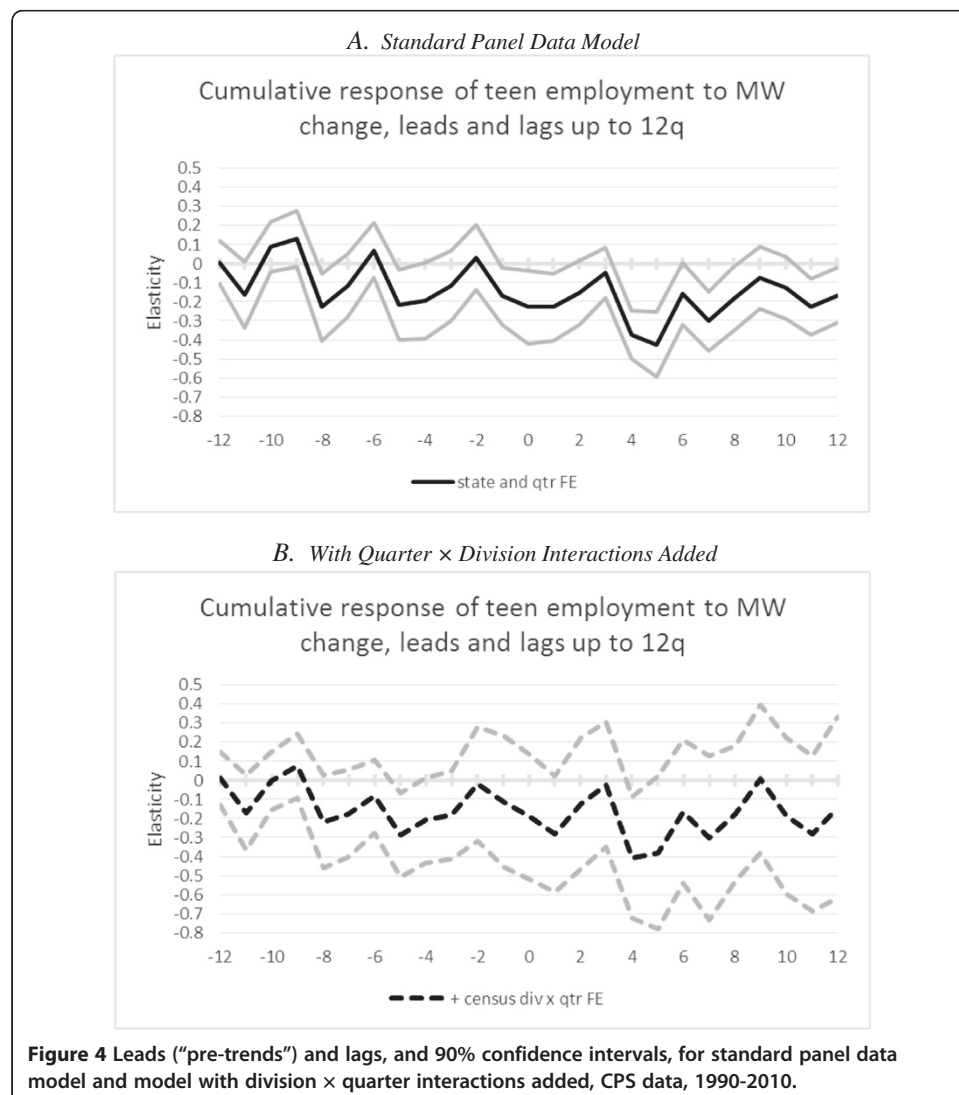**Figure 3 Leads ("pre-trends") and lags for alternative estimators, CPS data, 1990-2010.**
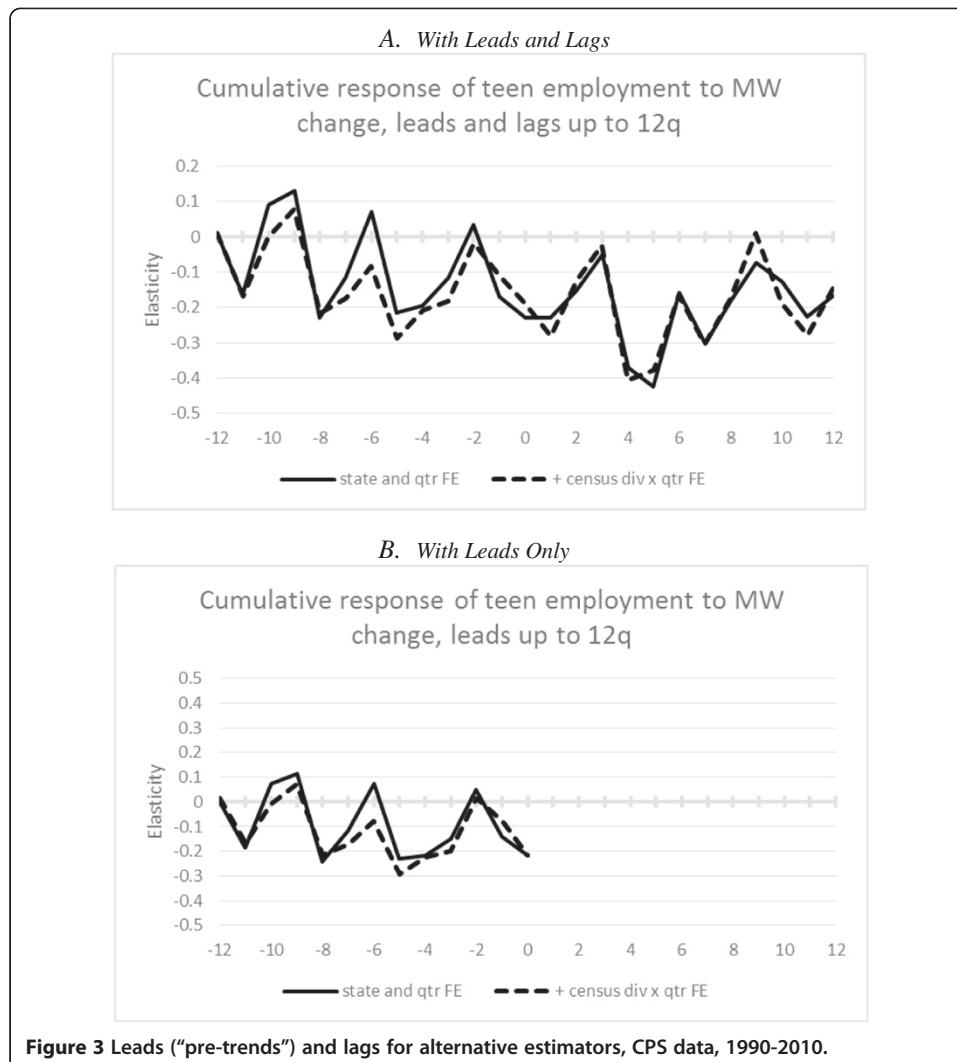
quarters. Looking first at the estimates for the specification including state and quarter fixed effects (the standard panel data model), we do not see evidence of a large accumulated negative effect in the period up to the minimum wage increase (i.e., the "pre-trends" that ADRZ claim plague the standard panel data estimator). The solid black does tend to be more negative than positive for the leading effects, but it returns to and goes above zero frequently. It is negative at a one-quarter lead, but that can be a real effect. The second point that emerges from this figure is that the model with the division × quarter interactions (the dashed line) actually looks worse, as it is persistently negative for most of the leads. Our conclusion, thus far, is that there is really no evidence to support ADRZ's claim that the standard panel data model is flawed because of strong pre-trends, in contrast to the model with spatial controls.

A second feature of Panel A of Figure 3 that is quite striking is that the estimates of the actual effects of minimum wages on teen employment are very similar whether or not the division × quarter interactions are included. Beginning at the contemporaneous effect or a quarter earlier, the solid and dashed lines by and large coincide. Thus, in this richer specification – allowing for leads and for longer lags – there is little evidence
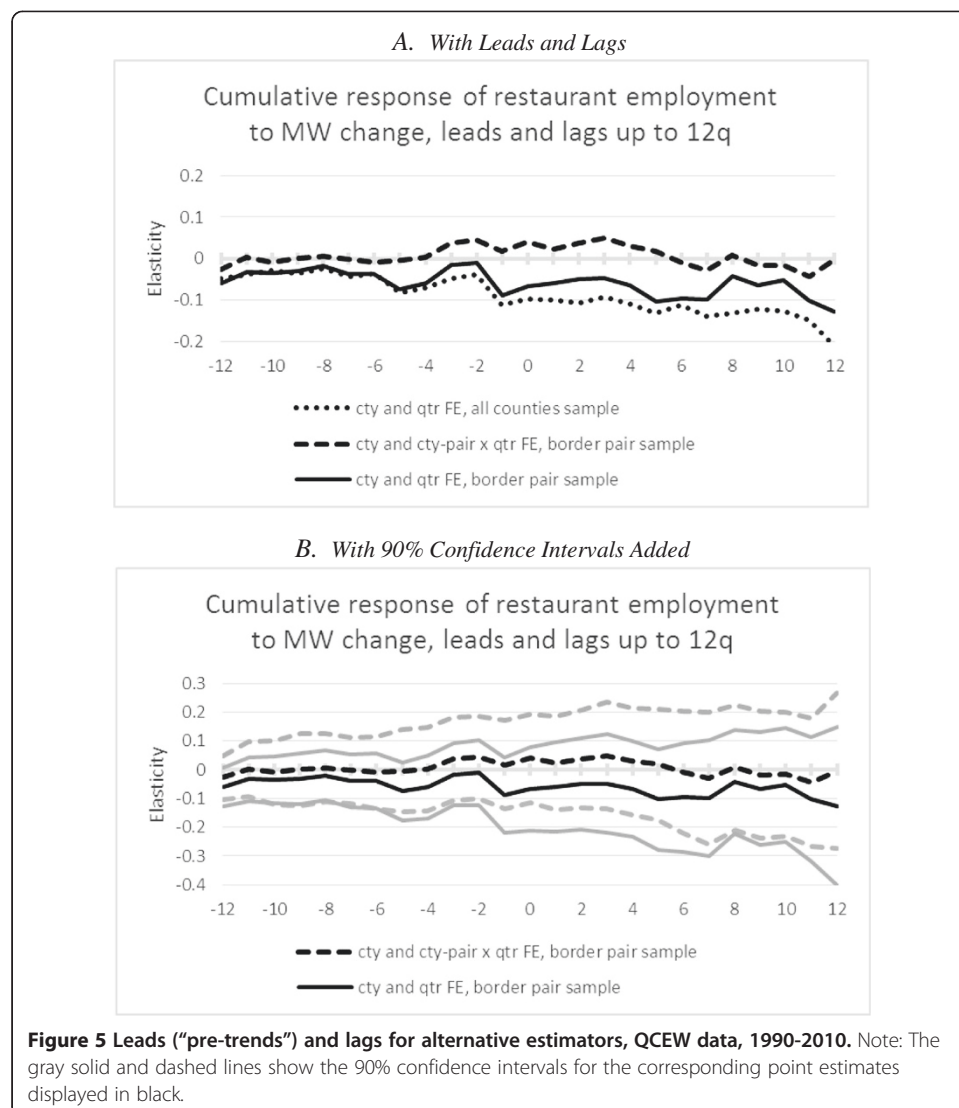
that the spatial heterogeneity controls that ADR used actually make much difference. One interpretation of this similarity of the results is that the inclusion of the leading effects controls for the kind of spatial heterogeneity that can make the comparison of the post-treatment to pre-treatment levels of the dependent variable a biased estimate of the treatment effect, yet without pinning identification on geographic proximity in the way that ADR do.[13] Panel B shows that the same conclusion emerges if we include only leads – which we do not advocate, but which is closer to the estimates ADRZ present.[14]

Moreover, our reading of the estimates in Panel A, using either the standard panel data model or adding the division × quarter interactions, is that there is a negative downward shift in teen employment that is most apparent about four to eight quarters after the minimum wage increase – consistent with a lagged effect discussed in much of the literature. What *is* different when the division × quarter interactions are added is that the estimates become much less precise. As can be seen by comparing Panels A and B of Figure 4, the confidence intervals for the model with quarter × division interactions added become much larger – sometimes as much as three times as large. In NSW, we emphasized that the research design advocated by ADR (and DLR) seemed to

Figure 4 Leads ("pre-trends") and lags, and 90% confidence intervals, for standard panel data model and model with division × quarter interactions added, CPS data, 1990-2010.

throw out a great deal of potentially valid identifying information, which apparently led to relying on invalid identifying information that biased their estimates. Here we see that in estimating richer models with leads and lags, there is also a large reduction in precision when geographically proximate controls are included. If there was clear evidence that this reduction in precision was necessary to remove an important source of bias from the estimates, the ADR/DLR estimators might be preferred. But as we have demonstrated in NSW and additionally in this paper (including in Figure 3), there is no such evidence.

Figure 5 reports a similar analysis for the QCEW data used in DLR. Two of the lines in Panel A contrast the estimates for the border pair sample with and without the interactions between the dummy variables for contiguous cross-border county pairs and quarter (these are the solid and dashed lines).[15] There is an indication – albeit modest – of lower employment beginning just prior to the minimum wage increase for the standard panel data model and continuing throughout the post-treatment period. There is also a slight indication of lower employment in the longer period leading up to the minimum

Figure 5 Leads ("pre-trends") and lags for alternative estimators, QCEW data, 1990-2010. Note: The gray solid and dashed lines show the 90% confidence intervals for the corresponding point estimates displayed in black.

wage increases, but there is nonetheless a distinct downward shift subsequently. Unlike for the CPS, in this case, there is not a substantial decline in precision from including the very large number of "spatial" controls in DLR's specification (Panel B).

Finally, in Panel A, we also show the results for all counties (the dotted line), as opposed to the border county subsample one needs to use to implement DLR's research design using contiguous cross-border counties. Given that we have shown that there is little justification for their research design, there is no reason to throw out information on all the other counties. The dotted line indicates clearer evidence of a downward employment shift after the minimum wage increase, although there is a slight tendency for lower employment before the minimum wage increase for this sample and specification as well.

Figure 6 rescales the lead and lag effects to show the cumulative effects of the minimum wage increase from one quarter prior to the increase (e.g., assuming that the one-quarter leading coefficient represents a real effect of the minimum wage) to 12 quarters after the minimum wage increase.[16] The idea is to estimate the minimum wage effect

Figure 6 Estimated cumulative minimum wage effects beginning one quarter prior to minimum wage increase, with 90% confidence intervals. Note: In Panels **A** and **B**, the gray solid and dashed lines show the 90% confidence intervals for the corresponding point estimates displayed in black.
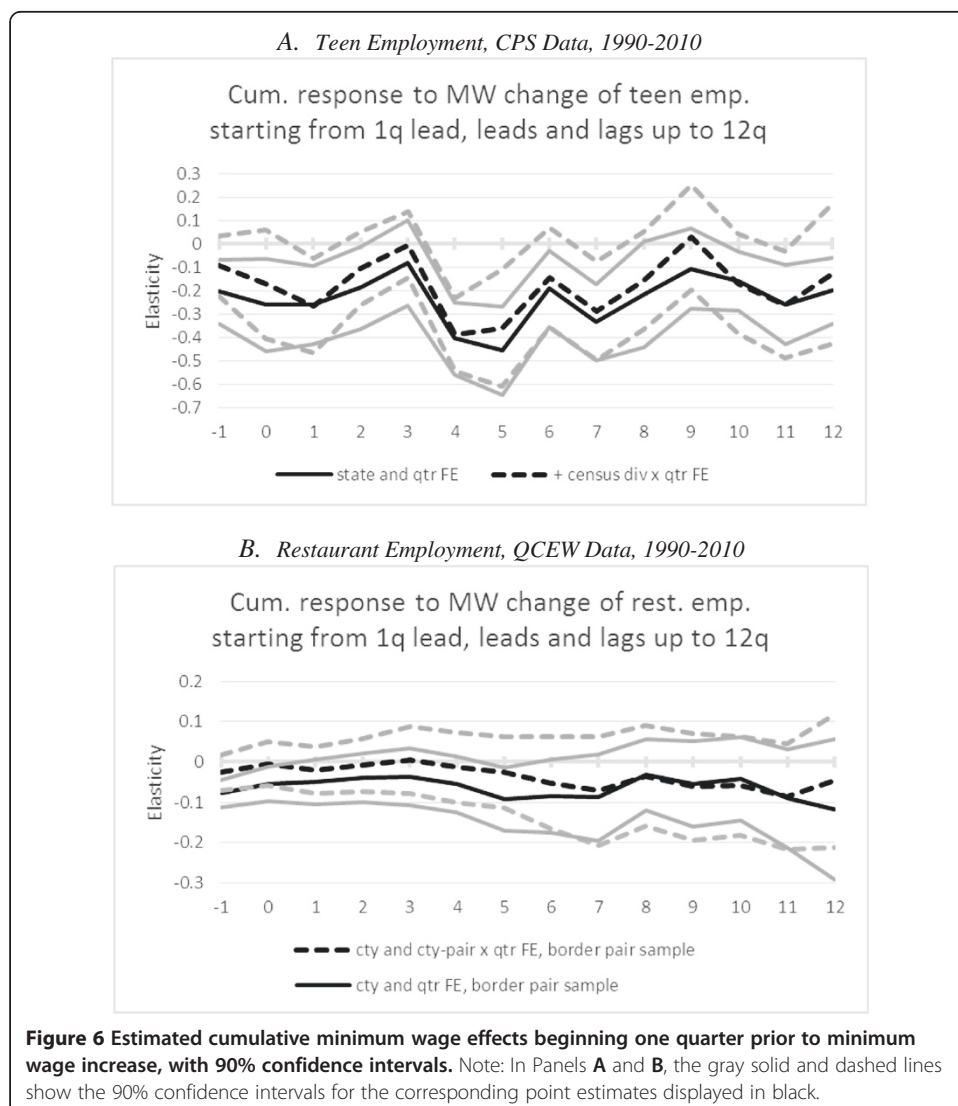
relative to employment in the preceding period, rather than relative to zero, to account for the possibility that there were employment changes prior to the minimum wage increase in some of the treatment states. For the state data, the contemporaneous elasticities are close to –0.2, building to a maximum of about –0.4 five quarters after the increase – a period around which the estimates are significantly different from zero. For the county-level analysis of restaurant employment using the border county subsample, the elasticities are between –0.05 and –0.1 after about five quarters and hold relatively steady through 12 quarters; these estimates are almost never significant.

Issue 3: Controlling for spatial heterogeneity using state-specific trends with different sample periods

A common robustness check in state-level panel data analyses is to include state-specific trends (and similarly for other units of analysis). In the analysis in ADR, adding linear state-specific trends to the specification eliminates the disemployment effects of minimum wages. However, in NSW, we showed that the inclusion of state-specific trends only eliminated the evidence of a negative effect of minimum wages under the restrictive specification that these trends are linear, and when one included endpoints of the sample period that included rather strong recessionary periods. We concluded that ADR's reliance on state-specific linear trends as part of their controls for "spatial heterogeneity" was likely invalid (which is why in the previous section we focused on the inclusion of the region × period controls), and established that their conclusions about the effects of introducing state-specific trends as local controls were very fragile.

ADRZ contest our analysis by doing one of two things to eliminate the influence of the problematic recessionary periods on the specifications with state-specific linear trends: they either drop the quarters that corresponded to recessionary quarters as indicated by NBER recession dates; or they include dummy variables interacting each of these spells with state dummy variables. The results are described in Table 4. As in NSW, we use data aggregated by state and quarter rather than micro-data, and do not have demographic controls. But the estimates that correspond in the two papers are

Table 4 The effects of the minimum wage on teen (16–19) employment, CPS data at state-by-quarter level, 1990 – 2011:Q2

|	Dependent variable: Log (employment/population)		
	State and quarter fixed effects	+ state-specific linear trends *and* division-quarter dummies	+ state-specific linear trends only
	(1)	(2)	(3)
A. Full sample, original specification (N = 4,386)			
Log(MW)	-.165*** (.041)	.009 (.058)	-.074 (.078)
B. Include state- and spell-specific recession dummy variables			
Log(MW)	-.169*** (.042)	.013 (.061)	-.065 (.080)
C. Leave out recessions (N = 3,672)			
Log(MW)	-.166*** (.041)	.026 (.061)	-.060 (.074)
D. Leave out 1990–1993, 1998–2011 (N = 2,754)			
Log(MW)	-.148** (.060)	…	-.229** (.095)

Notes: Estimates are weighted by teen population. The specifications include the state unemployment rate for adults and the relative size of youth population as controls. Standard errors are clustered at the state level. *Statistically significant at the .10 level; ** at the .05 level; *** at the .01 level.

very similar, so this is inconsequential. Focus first on columns (1) and (2), rows A-C, which correspond to ADRZ's Appendix Table B2. With just fixed state and quarter effects, in column (1), the estimated minimum wage elasticity is –0.17 for the entire sample period (row A). Doing the two things ADRZ suggest – leaving out recessions or including recessions-state interactions (rows B and C) – the estimates barely change relative to Panel A. The evidence they emphasize is in column (2), where they show that the estimated elasticity is near zero and does not change appreciably with their approaches to removing the effects of the recessions when they include the state-specific linear trends.

Note, however, that ADRZ also include the Census division-quarter interactions in these specifications. When we restrict attention to what happens to the specifications with only the linear trends added (column 3), the estimates are more negative, although not statistically significant. Still, they are not very sensitive to the two ways ADRZ propose – in rows B and C – to account for recessions.

As Figure 7 makes clear, however, the correspondence between the recession dates and when labor markets were in very weak territory is highly imperfect, as the recessions' impacts extended well beyond the official recession dates. Indeed when ADRZ drop *only* the recession periods, they *include* the periods after the recessions when labor market performance was even worse. Thus, their alternative specifications do not solve the endpoint bias problem, as the labor market was still very weak after the end of the early recession in 1991:Q1, as well as after the Great Recession's official ending date of 2009:Q2. This is why we dropped longer periods at the beginning and end of the sample period. As already noted, and as shown in row D of the table, when we do this (or any of the other things we do to get more robust estimates of the trends), we get strong negative disemployment effects.[17]

ADRZ also criticize another approach we use – using a Hodrick-Prescott (HP) filter to detrend the data. They suggest that this approach is inconsistent with our motivation "that cyclical downturns may be problematic to include" and hence suggest that "it is odd to hone in on the variation that the HP filter characterizes as business cycle
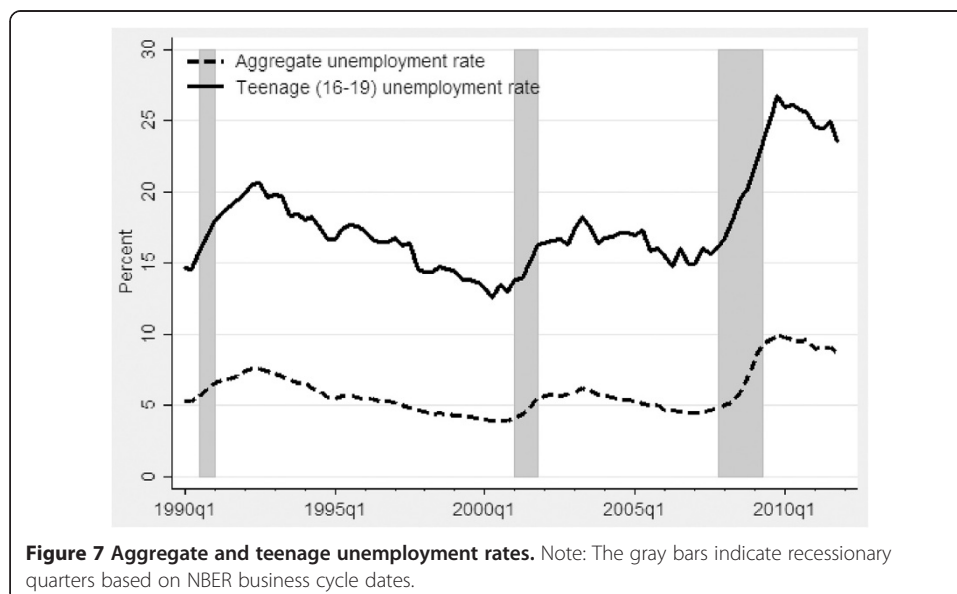
Figure 7 Aggregate and teenage unemployment rates. Note: The gray bars indicate recessionary quarters based on NBER business cycle dates.

variation" (p. 74). The problem is not the variation that occurs over the business cycle, per se. The problem we were emphasizing is that the business cycle can inappropriately influence the estimates of the linear trends when the endpoints include the effects of sharp downturns. In contrast, we do not want to discount the information on what happens to teen employment during recessions in relation to the minimum wage, but rather to avoid this information being down-weighted by fitting trends through these recessionary periods. Thus, the motivation for the filtering is to extract the trend in a manner that is less sensitive to the endpoint problem – which we do with the HP filter, as well as a more brute-force way of removing the trend using peak-to-peak comparisons and including non-linear state-specific trends.[18]

At the end of the day, ADRZ appear to concede, as our work established, that estimates that attempt to remove long-run trends are sensitive to parametric assumptions. However, this strikes us as a far cry from their original conclusions. For example, in the original paper ADR wrote: "These results indicate that estimates of minimum wage employment effects using the standard fixed-effects model of specification 1 are contaminated by heterogeneous employment patterns across states. Allowing for long-term differential state trends makes the employment estimates indistinguishable from zero" (2011, p. 220). Rather, they should have concluded, at most, that the estimates are sensitive to how one controls for these trends. Furthermore, nothing in this exchange undermines our main point that there are many sensible things one can do to avoid the problem of the business cycle influencing the estimates of long-run trends, and that the evidence from nearly all of these – except a couple of restrictive or inappropriate approaches – points to disemployment effects of minimum wages.

Issue 4: Placebo/falsification test for spatial heterogeneity

As one way of testing for spatial heterogeneity bias in the standard panel data model, DLR conduct a placebo or falsification test by estimating the "effects" of the cross-border minimum wage – which should not have a real effect but could have a spurious negative effect if the cross-border county is subject to the same negative shocks associated with minimum increases. They argue that their evidence establishes that the standard panel data estimator fails this test, exhibiting disemployment effects of minimum wages when there is no minimum wage variation.

Placebo tests can be a useful way to assess whether treatment effects are real or spurious, but in NSW, we argued that DLR's test was invalid because federal minimum wage changes across the border would also imply minimum wage changes in the state in question, so that the cross-border minimum wage assigned to placebo observations is contaminated with actual minimum wage variation. We therefore restricted the placebo sample not only to observations where the minimum wage was never above the federal minimum wage, but *also* limited the sample to when there was no federal minimum wage variation (and also examined a more-restrictive sample that included only county pairs with a minimum wage difference for at least one quarter). In these analyses, we find no statistical evidence of spurious disemployment effects of minimum wages.

ADRZ take issue with our critique of the falsification test in DLR. In particular, they state that we "misunderstood this entire exercise" and "the basic sources of statistical variation used in a fixed effects model" (p. 76). In fact, their argument is simply incorrect.

ADRZ note that if we take a sample of observations where the federal minimum wage always prevails, so $MW_{st}^S = MW_t^F$ for all t, and let S' denote the bordering state, then the regression model

$$E_{st} = \gamma MW_{st}^S + \delta MW_{st}^{S'} + D_s\theta + D_t\lambda + \varepsilon_{st} \tag{3}$$

is the same as

$$E_{st} = \gamma MW_t^F + \delta MW_{st}^{S'} + D_s\theta + D_t\lambda + \varepsilon_{st} \tag{4}$$

because $MW_{st}^S = MW_t^F$.

Clearly in equation (4), γ is unidentified since the federal minimum wage is perfectly collinear with the period dummy variables, and we get the same estimate of δ even if we drop MW_{st}^S (= MW_t^F) and just estimate

$$E_{st} = \delta MW_{st}^{S'} + D_s\theta + D_t\lambda + \varepsilon_{st}. \tag{5}$$

This is the equation DLR estimate, and in which, they argue, the estimate of δ provides a falsification or placebo test.

Based on the equations above, ADRZ argue that federal minimum wage variation is irrelevant and cannot be contaminating the falsification experiment.[19] However, this is not true. MW_t^F is perfectly collinear with the period fixed effects. But $MW_{st}^{S'}$ in equation (5) varies with the federal minimum wage in a way that is *not* perfectly correlated with the period fixed effects, because whether the federal minimum wage variation changes the cross-border minimum wage depends on whether the state or federal minimum wage is binding. Thus, federal minimum variation is not swept out by the period fixed effects, and therefore the cross-border minimum wage variation will be correlated with the actual state minimum wage variation.

Another way to see this is to note that the regression they estimate for their placebo test is

$$E_{st} = \delta\left(MW_t^F \cdot I\{MW_{st}^{S'} = MW_t^F\} + MW_{st}^{S'} \cdot I\{MW_{st}^{S'} > MW_t^F\}\right) + D_s\theta + D_t\lambda + \varepsilon_{st}, \tag{6}$$

where $I\{\cdot\}$ is the indicator function. That is, there is a single minimum wage coefficient that is constrained to be the same whether or not the minimum wage variation is coming from the federal minimum wage or the state minimum wage across the border. Clearly the federal variation can play a role here because the federal minimum wage is multiplied by a dummy that is sometimes one and sometimes zero, breaking the perfect collinearity with the time fixed effects.

An obvious way to verify that federal minimum wage variation plays a role is to vary artificially the federal minimum wage, always being careful to keep track of what this does to state minimum wages in the cross-border state – nothing if it stays below the state minimum, but changing it if the federal minimum wage is binding. If we do this, and the estimated minimum wage effect in equation (5) changes, then clearly the federal minimum wage variation plays a role. As documented in Table 5, columns (2) and (3), this does in fact change the estimated effect of the cross-border minimum wage in DLR's placebo sample and regression, relative to the estimates in column (1), which come from their paper.

Table 5 The effects of the minimum wage on log restaurant employment, "falsification tests," county-level QCEW data

	DLR estimates	False variation in federal minimum wage		Correct Maryland minimum wage error		
	DLR data and estimates: 1990:Q1-2006:Q2 (Table B1, specification 2)	Federal minimum lower by 5 cents (4.20/4.70/5.10 instead of 4.25/4.75/5.15 beginning in 91:Q2/96:Q4/97:Q3)	Federal minimum wage always = $3.35	Same as column (1), but with corrected Maryland minimum wage	Sample restricted to 1998:Q3-2006:Q2 (period with no federal MW changes), with corrected Maryland minimum wage	Sample restricted to 1998:Q3-2006:Q2, county pairs with minimum wage difference for at least one quarter, with corrected Maryland minimum wage
	(1)	(2)	(3)	(4)	(5)	(6)
Actual MW sample:						
Log(MW)	−0.208 (0.150)	…	…	−0.114 (0.155)	−0.198[**] (0.079)	−0.174[**] (0.085)
Placebo MW sample:						
Log(MW)	−0.123 (0.158)	−0.111 (0.157)	−0.099[**] (0.042)	−0.125 (0.134)	−0.088 (0.062)	0.035 (0.100)

Notes: All specifications include period and county fixed effects, and controls for population and private-sector employment. Standard errors are clustered at the state level. *Statistically significant at the .10 level; ** at the .05 level; *** at the .01 level. (See NSW, 2014, for more details).

As one additional point regarding DLR's placebo test, their dataset includes an error that is influential in this analysis. (It was not influential in other analyses.) They have a higher minimum wage in Maryland in the first six months of 2006 – $6.15, instead of the $5.15 federal minimum wage that actually prevailed. The effects of this error are illustrated in the final three columns of Table 5. When the corrected data are used, but with their sample (column (4)), the estimates are very similar for the actual minimum wage sample and the placebo sample (–0.114 and –0.125). Thus, in this case there is perhaps not even a basis for a placebo test since the initial estimate is quite small. Nonetheless, with the corrected data, the estimates for columns (5) and (6) – which avoid the problem with their placebo test being invalid – parallel those in our paper. The column (5) estimates are –0.198 (0.079) for the actual sample and –0.088 (0.062) for the placebo sample, and the column (6) estimates are –0.174 (0.085) and 0.035 (0.100), respectively. Thus, with the corrected data, we get a clear negative estimate for the actual data and no effect for the placebo sample – exactly what should happen if the negative estimated employment effect is *not* driven by a spurious correlation of minimum wage increases and negative shocks that are common across the border, but instead represents a real effect of the minimum wage.

Why might focusing on local variation produce biased estimates of minimum wage effects?

ADRZ assert that the basis of our critique was that their spatial controls "discard too much variation to find any significant effects" (p. 22). This assertion is incorrect; in NSW, we did not focus on the loss of precision. Rather, after testing the restrictions entailed by their local estimators, we concluded that the evidence indicated that DLR (and ADR) were arbitrarily throwing away lots of valid identifying information and potentially focusing on variation that generated biased estimates. It is the case, however, that in longer-term dynamic models, loss of precision is an issue as well, as we note in this paper.

Nonetheless, in NSW, we did not offer an explanation of why DLR (or ADR's) estimates based on local comparisons generate biased estimates of minimum wage effects, in particular estimates biased upward toward zero. There is actually a relatively simple potential explanation for this bias. Baskaya and Rubinstein (2012) find that when the interaction between the federal minimum wage variation and the propensity for the federal minimum wage to bind in a state is used as an instrumental variable for the prevailing state minimum wage, stronger negative disemployment effects result compared to the standard panel data estimator.

A natural interpretation of this evidence is that because state minimum wages tend to be more similar to those in nearby states than in states farther away, when the identifying variation is restricted to same-division or cross-border states, much of the identifying information from federal minimum wage variation is eliminated. Instead, identification comes more from state variation that is unrelated to federal variation, which is more likely to be associated with positive endogeneity bias that comes from policymakers tending to raise the minimum wage when the labor market is stronger. Baskaya and Rubinstein present some indirect evidence consistent with this, finding that in states where the minimum wage is generally above the federal minimum wage, the minimum wage moves pro-cyclically with about a one-year lag with respect to the

unemployment rate, whereas minimum wage variation in states where the federal minimum wages binds is not correlated with the lagged unemployment rate. Of course we are more concerned about endogeneity with respect to low-skill labor markets per se.

More generally, restricting the usable variation to narrow regions does not necessarily produce less biased estimates of minimum wage effects. In particular, including region × period interactions can wipe out the across-region variation in minimum wages attributable to influences like changes in inequality, union strength, or the federal minimum wage that are largely exogenous with respect to shocks to local low-skill labor markets and which differ more across regions than within regions.[20] In contrast, we might ask why minimum wage changes differ in similar, nearby states in the same period. Our conjecture is that the within-region and within-period variation is driven more by unmeasured variation in low-skill labor market conditions to which policymakers respond by raising minimum wages when these conditions are strong, generating positive bias in the estimated employment effect of minimum wages. The findings in Baskaya and Rubinstein (2012) are consistent with this conjecture.

A useful analogy comes from Griliches' (1979) seminal work on twin or sibling estimates of the economic returns to schooling. The simplest intuition is that if we include family fixed effects, or equivalently look only at within-family variation in schooling and wages, then bias from omitted unobservables at the family level is reduced. Griliches pointed out, however, that whether or not this is true depends on what generates variation within versus across families. In particular, if family influences or "background" common to both siblings or twins are relatively important in determining schooling, then the remaining within-family differences can be more reflective of ability differences to which schooling responds, in which case the within-family estimate of the return to schooling can be more biased than estimates using across-family variation. This parallels the case above when the within-region and period variation in minimum wages is more closely related to unmeasured variation in local low-skill labor markets than is the across-region (and period) variation. In contrast, if within-family schooling differences are less driven by the common influences on siblings or twins, then more of the within-family differences are determined by factors other than ability, and the within-family estimate will be less biased. This corresponds to the scenario ADR and DLR assume, in which the within-region and period variation in minimum wages is more exogenous to local labor market conditions. However, we have explained that the available evidence suggests that the former scenario may be more plausible.

ADRZ assert, with respect to their inclusion of period-region fixed effects, that, "there are only two acceptable reasons to avoid controlling for this heterogeneity. (1) The inclusion of the controls substantially reduces statistical power. (2) The treatment affects the control variables themselves, such as through spillover effects on neighboring areas" (p. 36). But as Griliches' work demonstrates – and the point has been echoed repeatedly in research with panel data where the issue is isomorphic to DLR's saturated models that focus on local variation – this statement is simply incorrect. Controlling for heterogeneity changes the identifying variation and can, under some circumstances, exacerbate other biases. And in this particular case, the upward endogeneity bias that we might expect in estimating the effects of minimum wages on employment is more likely to emerge with the inclusion of local controls.

Conclusions

Our original paper (NSW, 2014) faulted two previous analyses by authors of ADRZ implementing research designs that change the comparison groups used in estimating the employment effects of minimum wages (ADR, 2011; DLR, 2010). Focusing on the key innovation in these papers – the use of geographically proximate areas as local controls – we concluded that while these research designs have some intuitive appeal a priori, the key identifying assumption underlying them – which generated the finding of no disemployment effects – was not supported by the data. In addition, when the data were used to pick out the best control regions for areas treated by a higher minimum wage, the standard disemployment effects were confirmed for teenagers using CPS data. The evidence for restaurant employment using QCEW data remains more ambiguous. Thus, our paper substantially undermined the contentions in ADR and DLR that essentially most of the research literature preceding their work, relying on conventional panel data estimators with fixed period and area (usually state) effects, used flawed comparison areas that generated spurious evidence of disemployment effects.

ADRZ are equally sweeping in their criticism of our evaluation, presenting a litany of criticisms of both our analyses and results and concluding that the findings in ADR and DLR stand. In this paper, we have attempted to highlight the main issues under debate regarding the selection of comparison areas or groups and demonstrate that the criticisms that ADRZ level at our analysis of these issues are unfounded. Indeed, we think it more likely that the restricted comparison groups they use result in estimates of minimum wage effects that are biased toward finding no disemployment effect. Finally, in a richer specification that includes leads and lags of minimum wages, prompted by specifications that ADRZ report, it is not even clear that the spatial heterogeneity controls that ADR used have much effect on the minimum wage effects estimated from the standard panel data model with fixed time and state effects.

At the end of the day, then, we end up where we started. We see the evidence as still pointing to disemployment effects for low-skilled workers from raising the minimum wage, with elasticities that are often around –0.2 for the teenagers on whom we focus. This evidence continues to be consistent with the comprehensive research literature reviewed in Neumark and Wascher (2007).

We are not under the illusion that our assessment of ADRZ's paper will settle the issue for all parties. The minimum wage-employment debate is contentious, and there is a continuing flow of new work that introduces new ideas or approaches pertaining to this debate, focusing in part on the same issue of the appropriate comparison groups for estimating minimum wage effects (e.g., Meer and West 2013; Aaronson et al., 2013).[21]

Indeed, given the potential endogeneity of minimum wage policy, we think it is important for researchers to continue their efforts to obtain more compelling identification of the effects of minimum wages. This is indeed the goal pursued by ADR and DLR, which in itself is commendable, even if the evidence indicates that their approach generates minimum-wage employment effects that are biased toward zero. We think the best study in this vein is the one by Baskaya and Rubinstein (2012), which uses an instrumental variable that relies on federally-induced minimum wage variation that is exogenous to the state and finds stronger evidence of disemployment effects, with elasticities often in the range of –0.4 or larger. In our view, then, the most recent evidence that merits serious consideration challenges the consensus view from the other side –

suggesting that the estimated disemployment effects of minimum wages on low-skill workers are substantially stronger than indicated by previous estimated elasticities in the range of −0.1 to −0.2. But this, too, will surely not be the last word.

Endnotes

[1]We focus to some extent on the results for teenagers using the CPS for three reasons. First, most of the debate in the literature is about estimated employment effects for low-skilled groups defined by age or other demographic characteristics, generally using the CPS. Second, for a number of reasons, predictions of disemployment effects are less clear for a single industry – including the restaurant industry where many workers are tipped. And third, the differences in the alternative estimates are most evident for the CPS teen results, which is where the disemployment effects predicted by the neoclassical model are sharpest.

[2]Because we do not rehash all of the details of the analyses from the prior papers, readers may find it useful to refer to NSW and ADRZ for a fuller understanding of the material we cover.

[3]Although the discussion here is in the context of the state-level analysis, the same argument applies to the county-level analysis.

[4]By "treatment" we mean a case where there is a minimum wage increase in a state that can be compared against non-treated states; the treatment is the minimum wage increase.

[5]It is not clear whether 0.938 or 0.925 is a more accurate characterization of the relative weights, although they are so close it hardly matters. Of course, neither is informative about the *distribution* of the relative weights across treatments, which is why we summarized the results in distributional terms in NSW.

[6]ADRZ object to our "using residuals from an OLS panel regression as the matching variable in a synthetic control study" (p. 65). However, we also present results from a synthetic control analysis that matches on various forms of the dependent variable, as well as one that matches on residuals from a specification that restricts the minimum wage coefficient to be zero (what ADRZ argue is the actual effect of minimum wages on employment). As we noted in our earlier paper, these alternative matching algorithms yielded very similar answers.

Allegretto et al. (2013b) also argue that the approach of matching on residuals is wrong because of "confusion between estimated and true residuals. By construction, estimated OLS residuals are uncorrelated with all regressors, including the minimum wage" (p. 28). Thus, they argue, the residuals are uninformative because they are "mean-zero errors that are uncorrelated with the minimum wage" (p. 28). It is of course true that the contemporaneous least-squares residuals are uncorrelated with the regressors by construction. But the matching is on *lagged* residuals, which are *not* uncorrelated by construction.

[7]One notable exception is for the West North Central states, where the weight in column (3) is 0.098, vs. 0.009 in column (8). But as shown in NSW (2014, Table 2), if we run the standard panel data model for this division, we get a standard disemployment elasticity of about −0.19, which is statistically significant. This kind of evidence for the West North Central states in NSW led to the conclusion that "(a) in most cases, there is little rationale for ADR's choice to focus only on the within-division variation to identify minimum wage effects; and (b) when there *is* a good rationale for doing this, the evidence shows negative and statistically significant effects of minimum

wages on teen employment, with elasticities that are in or near the −0.1 to −0.2 range"
(p. 627).

[8]We use the centroids of each state (county) in calculating the distances between them, and we non-parametrically estimate the average synthetic control weight by distance from treatment state (county) using locally-weighted regression (lowess command in Stata). In implementing lowess, we vary the smoothing method (running-line least squares, which is the default, or running-mean) and bandwidth used (the default 0.8 or a narrower 0.4).

[9]The right-hand tails reflect distances from Hawaii and Alaska to select states, and hence are not very reliable.

[10]ADRZ have to date declined to provide their data and code with which we could compare ours.

[11]Additional discussion of our conclusions regarding state-specific linear trends appears below.

[12]A real effect can also presumably arise with a short lead, owing to employers' responses to soon-to-be-implemented minimum wage increases.

[13]As ADRZ point out, another way to control for prior changes in the model with state/county and period fixed effects is to include lagged dependent variables (LDVs). Here too, we came up with results that are quite different from what ADRZ present in columns (3)-(4) of their Table 4; in particular, we still find significant negative employment effects for teens in the CPS data. Using quarterly CPS data from 1990 to 2010, our estimates show a teen employment elasticity that ranges from −0.057 to −0.108 (one-quarter LDV) or from −0.034 to −0.071 (one- to four-quarter LDV), with all the estimates statistically significant. This contrasts with their estimates, which range from −0.004 to −0.076 (one-year LDV using annual CPS data from 1990 to 2012), with the latter estimate statistically significant only at the 10% level.

[14]A footnote in NSW also reported results using bordering state pairs instead of Census divisions to construct controls, more closely paralleling the county-level research design in DLR. This approach tends to give stronger evidence of disemployment effects than ADR's research design based on Census divisions. However, there is a bit more evidence of negative pre-trends for this specification, although nothing as severe as what ADRZ report; and there is a sharper negative shift after the minimum wage takes effect. (See Additional file 1: Figure A1.)

[15]These QCEW estimates are for the specifications that include the private-sector employment control, and both use the contiguous border county pair sample originally used in DLR. Note that in ADRZ's Figure 6, the sample used in what is presented as estimates for the canonical model is different – it is the all counties sample, which is not directly comparable to the contiguous border county pair sample used for the local controls model.

[16]To do this, we subtract out the cumulative effect through two quarters prior to the minimum wage increase.

[17]ADRZ obfuscate the issue by questioning whether the periods we omit constitute "recessions" (their footnote 49). The NBER recession dates do not include the entire periods of 1990–1993 and 2008–2011, and we never stated that we were explicitly leaving out recessions based on their formal start and stop dates. The data in Figure 7 clearly point to labor markets in which both aggregate and teen unemployment were unusually high in these periods. We are also well aware that there was a recession in 2001, but we explicitly discussed the problem of recessionary periods coming at the beginning or end of periods over which trends are estimated.

[18]Despite ADRZ's claim (their footnote 50) to the contrary, the HP filter is sometimes used to detrend panel data in applied microeconomics papers, especially when the panels consist of aggregate data for geographic areas across time. For an example using data for U.S. states, see Ionides et al. (2013), who make a similar point to ours about using nonlinear detrending techniques. In addition, the reference entry for the STATA command "tsfilter hp" explicitly indicates that it is designed to be compatible with panel data and that the filtering is done separately on each panel.

[19]They do not quite say it this way. Rather, they write (of minimum wages): "They are changing, however, in exactly the same way in all counties in the placebo sample, since they all pay the federal minimum and are fully correlated with time effects. In other words, there is *zero* cross-sectional variation in minimum wages in the sample" (p. 76). The language is a little imprecise, but their argument must be that the federal minimum wage variation is completely absorbed in the period fixed effects, and therefore there is no remaining minimum wage variation in the placebo sample to be "predicted" by the federal minimum wage.

[20]ADRZ document, in their Table 1, that high minimum wage states tend to differ from low minimum wage states in terms of unionization, voting patterns, changes in inequality, and the business cycle. As we argued in NSW (2014), differences in the business cycle are captured in the aggregate labor market controls included in the models.

[21]On the Meer and West study, see Dube's (2013) criticism, and their reply at http://econweb.tamu.edu/jmeer/Meer_West_MinimumWage_Appendix.pdf (viewed February 16, 2014).

Additional file

Additional file 1: Figure A1. Leads ("pre-trends") and lags for alternative estimators, adding paired-state border design, CPS Data, 1990-2010.

Author details
[1]UCI, NBER, and IZA, 3151 Social Science Plaza, Irvine, CA 92697, USA. [2]Harvard University, Cambridge, MA, USA. [3]Board of Governors of the Federal Reserve System, Washington, DC, USA.

References
Aaronson D, French E, Sorkin I (2013) Firm dynamics and the minimum wage: A putty-clay approach. Federal Reserve Bank of Chicago Working Paper No. 2013-26, Chicago, IL, USA
Abadie A, Diamond A, Hainmueller J (2010) Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. J Am Stat Assoc 105(490):493–505
Allegretto SA, Dube A, Reich M (2011) Do minimum wages really reduce teen employment? Accounting for heterogeneity and selectivity in state panel data. Ind Relat 50(2):205–240
Allegretto SA, Dube A, Reich M, Zipperer B (2013a) Credible research designs for minimum wage studies. IZA Discussion Paper No. 7638. Bonn, Germany
Allegretto S, Dube A, Reich M, Zipperer B (2013b) Credible research designs for minimum wage studies. Working paper, Institute for Research on Labor and Employment, UC Berkeley, June 11, 2013, http://www.irle.berkeley.edu/workingpapers/148-13.pdf (viewed August 30, 2013)

Baskaya YS, Rubinstein Y (2012) Using federal minimum wage effects to identify the impact of minimum wages on employment and earnings across U.S. states. Central Bank of Turkey, Turkey, Unpublished paper

Dube A (2011) Review of Minimum Wages by David Neumark and William Wascher. J Econ Lit 49(3):762–766

Dube A (2013) Minimum wages and aggregate job growth: Causal effect or statistical artifact? IZA Discussion Paper No. 7674. Bonn, Germany

Dube A, Lester TW, Reich M (2010) Minimum wage effects across state borders: Estimates using contiguous counties. Rev Econ Stat 92(4):945–964

Griliches Z (1979) Sibling models and data in economics: Beginnings of a survey. J Pol Econ 87(5,Pt2):S37–S64

Ionides EL, Wang Z, Tapia Granados JA (2013) Macroeconomic effects on mortality revealed by panel analysis with nonlinear trends. Ann App Stat 7(3):1362–1385

Meer J, West J (2013) Effects of the minimum wage on employment dynamics. NBER Working Paper No. 19262, Cambridge, MA, USA

Neumark D, Wascher W (2007) Minimum wages and employment. Foundations Trends Microecon 3(1-2):1–182

Neumark D, Salas JMI, Wascher W (2014) Revisiting the minimum wage-employment debate: Throwing out the baby with the bathwater? Ind Labor Relat Rev 67(3):608–648