American Economic Journal: Economic Policy 2021, 13(2): 58–99
https://doi.org/10.1257/pol.20160613

Unemployment Insurance Generosity and Aggregate Employment†

By Christopher Boone, Arindrajit Dube, Lucas Goodman, and Ethan Kaplan*

This paper examines the impact of unemployment insurance (UI) on aggregate employment by exploiting cross-state variation in the maximum benefit duration during the Great Recession. Comparing adjacent counties located in neighboring states, there is no statistically significant impact of increasing UI generosity on aggregate employment. Point estimates are uniformly small in magnitude, and the most precise estimates rule out employment-to-population ratio reductions in excess of 0.35 percentage points from the UI extension. The results contrast with the negative effects implied by most micro-level labor supply studies and are consistent with both job rationing and aggregate demand channels. (*JEL* E24, E32, J22, J23, J65)

During the Great Recession, existing law and new acts of Congress led to the most dramatic expansion in the duration of unemployment insurance (UI) benefits in US history.[1] In most states, eligible job losers saw their maximum benefit duration rise from the usual 26 weeks to 99 weeks. Continuously from November 2009 through March 2012, the maximum benefit duration exceeded 90 weeks when averaged across states, except for a few small lapses in congressional authorization. In comparison, during a previous spell of extended benefits in response to the 2001 recession, this average rarely exceeded 40 weeks (Farber and Valletta 2015).

This unprecedented UI expansion—and its variation across states in magnitude and timing—provides a unique opportunity to study the aggregate employment effects of UI benefit duration. In this paper, we examine the effect of the

* Boone: University of Massachusetts Amherst, 412 N. Pleseant St., Amherst, MA 01002 (email: cboone@umass.edu); Dube: University of Massachusetts Amherst, 1030 Thompson Hall, Amherst, MA 01003 (email: adube@econs.umass.edu); Goodman: Office of Tax Analysis, U.S. Treasury Department, 1500 Pennsylvania Ave NW, Washington, DC 20220 (email: lucas.goodman@treasury.gov); Kaplan: Department of Economics, 3114 Tydings Hall, University of Maryland at College Park, College Park, MD 20743 (email: kaplan@econ.umd.edu). Matthew Shapiro was coeditor for this article. We thank Gabriel Chodorow-Reich, Thomas Hegland, Ioana Marinescu, Jesse Rothstein, and three anonymous referees for helpful comments. We additionally thank seminar participants at a 2017 AEA session, Auburn University, and George Mason University. Dube and Kaplan acknowledge financial support from the Institute for New Economic Thinking. Kaplan also acknowledges financial support from the Washington Center for Equitable Growth. We wish to thank Doruk Cengiz, Bryan Hardy, and Yuting Huang for excellent research assistance. The views expressed in this report are the authors' and should not be interpreted as those of Treasury.

† Go to https://doi.org/10.1257/pol.20160613 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

[1] The second-largest increase provided UI duration of 65 weeks in 1975 following the passage of the Special Unemployment Insurance Extension Act.

maximum benefit duration on aggregate employment during the Great Recession using state-level expansions and contractions in UI generosity. We use county-level monthly employment data from late 2007 through the end of 2014. We provide transparent evidence on employment dynamics around sharp and durable changes in UI benefits across state boundaries. Our paper relates closely to two influential papers by Hagedorn et al. (2019) (henceforth HKMM) and Hagedorn, Manovskii, and Mitman (2016) (henceforth HMM), and we employ similar empirical strategies. While both HKMM and HMM find substantial negative effects of UI on employment, we find effects that are close to and statistically indistinguishable from zero. In this paper we additionally provide a detailed reconciliation of the differences between our findings and those of the two Hagedorn papers.

A large body of research has studied the effect of UI duration on the labor supply and job search behavior of individuals. This includes evidence from the Great Recession in the United States (Rothstein 2011; Farber, Rothstein, and Valletta 2015; Farber and Valletta 2015; Johnston and Mas 2018). Of course, the overall macro effects of benefit extensions on aggregate employment may be quite different from the micro-based estimates. If jobs are rationed, the decreased search from increased UI generosity during downturns may have only limited effects on aggregate employment due to increased labor market tightness—implying a less negative macro effect than micro effect (Michaillat 2012; Lalive, Landais, and Zweimüller 2015; Landais, Michaillat and Saez 2018). Alternatively, the overall impact on employment could be more negative than predicted by the micro effects if an increase in reservation wages causes firms to reduce vacancies (Mitman and Rabinovich 2015). Finally, Keynesian theory predicts that UI provision during recessions could help boost employment via an aggregate demand channel, which, if large enough, could even lead to an overall positive macro effect (Kekre 2019).

Unfortunately, a small set of recent empirical papers has delivered a mixed verdict on the size of the macro effect of the policy (Amaral and Ice 2014; Hagedorn, Manovskii, and Mitman 2016; Johnston and Mas 2018; Chodorow-Reich, Coglianese, and Karabarbounis 2019; Hagedorn et al. 2019; Dieterle, Bartalotti, and Brummet 2020). While HKMM and HMM estimate large negative effects on employment, other papers using alternative empirical strategies find small and statistically insignificant aggregate effects of the UI benefit expansion, including Chodorow-Reich, Coglianese, and Karabarbounis (2019) and Dieterle, Bartalotti, and Brummet (2020). Chodorow-Reich, Coglianese, and Karabarbounis (2019) devise an innovative approach to estimating the employment effects of UI by comparing states with the same unemployment rate but different UI duration due to differences in real-time measurement error of the unemployment rate. Dieterle, Bartalotti, and Brummet (2020) investigate the shortcomings of empirical approaches that rely on comparing outcomes across state boundaries. They also develop a regression-discontinuity-based method that more heavily uses variation from counties with populations that live closer to each other. We see these approaches as complementary to ours. Below, we elaborate on the differences in empirical designs across these papers, along with our contribution.

Analyzing the effect of UI expansions during the Great Recession is difficult due to a reverse causality problem; under the rules of the expansion, states with a higher unemployment rate were entitled to longer benefits. Our two main empirical strategies endeavor to overcome this "mechanical endogeneity" problem by making use of a border-county-pair (hereafter BCP) design where we compare employment in neighboring counties located on opposite sides of state borders. Here we follow the recent UI literature (beginning with HKMM) that uses a BCP design to estimate the impacts of UI extensions (Amaral and Ice 2014; Hagedorn, Manovskii, and Mitman 2016; Dieterle, Bartalotti, and Brummet 2020). The BCP design relies on the fact that counties in neighboring states are geographically contiguous and (as we show) economically similar but often experienced different UI duration during our sample period due to differences in both state policy and state unemployment rates. We show that the BCP strategy substantially reduces the endogeneity problem, mitigating negative preexisting employment trends in counties that subsequently experienced greater expansions in maximum benefit duration.[2]

We use two empirical strategies. Our first empirical strategy compares employment outcomes within county pairs using monthly data on employment and UI duration from the end of 2007 through the end of 2014. This full sample panel specification makes use of all variation in state-level UI duration over the entire period, including UI expansions that are triggered by an increase in state-level unemployment. We also provide a second "event study"–type strategy that uses variation induced by national-level policy changes to instrument for the changes in state-level UI duration. In particular, we use the November 2008 expansion and the December 2013 expiration of the Emergency Unemployment Compensation (EUC) program. These national-level policy changes eliminate endogenous state-level triggers as a source of variation and thus are more plausibly exogenous to employment changes between neighboring counties than variation resulting from the movements in state-level unemployment rates. At the same time, the bite of the policy differed across state borders, which allows us to use the BCP strategy in conjunction with the event study approach. We show changes in aggregate employment during the 12 months before and after these expansion and expiration events; we also combine the data for both events to produce a pooled estimate. While the event study has less statistical power, it has the advantage of being a cleaner identification strategy.

We additionally provide a data-driven refinement of the BCP approach where we trim county pairs with a high level of presample mean-squared error in employment between the counties. This restricts the analysis to the set of county pairs that are more alike and hence have lower error variance. While the point estimates from the refined and trimmed samples are similar, the reduction in noise makes estimates from the trimmed sample more precise—allowing us to rule out even quite modest negative employment effects.

Our main results are as follows. We find no evidence that UI benefit extensions substantially affected county-level employment. For the full sample regressions, our point estimates for the effect of expanding maximum benefit duration from

[2] An early example of the border-county-pair strategy was by Dube, Lester, and Reich (2010), who used it to study minimum wage policies that changed discontinuously at state borders. Note that the same problem of mechanical endogeneity does not arise when studying the effects of the minimum wage, as statutory wage rates are not directly tied to measures of local employment or unemployment.

26 to 99 weeks range from an increase of 0.18 to 0.43 percentage points in the employment-to-population (EPOP) ratio. These estimates are not significantly different than zero; using the more precise estimates from the trimmed sample, one can rule out an effect more negative than −0.35 percentage points at the 95 percent confidence level. By comparison, the total change in EPOP over the course of the Great Recession was about −3 percentage points in our sample. Our event study estimates yield similar but somewhat less precise results, showing that employment remained stable prior to and following treatment. We additionally compare our macro employment estimates of UI to the micro estimates based on individual labor supply from prior studies. Most of the estimates suggest a positive gap between the macro and micro estimates, with a possible explanation being that UI increases labor market tightness. At the same time, some of our point estimates indicate an overall positive effect of UI on employment (though all of our 95 percent confidence intervals include zero). A positive aggregate employment effect of UI could not be explained by the tightness channel alone but could result from an additional aggregate demand effect (Kekre 2019).

Our findings differ greatly from both HKMM and HMM, who estimate large negative effects of UI on aggregate employment.[3] In Section V, we provide a reconciliation of our results with those from HKMM and HMM, fully decomposing the two sets of results into constituent factors. With respect to HKMM, we find that three factors account for almost all of the gap in the estimates. First, HKMM quasi-forward difference their dependent variable. This imposes a strict parametric assumption on the dynamic effects of UI, which we show is unwarranted, and can produce an estimate that has the incorrect sign and an amplified magnitude. Second, we make use of the higher quality administrative data from the Quarterly Census on Employment and Wages (QCEW) rather than the Local Area Unemployment Statistics (LAUS), which is partially imputed using confounding state-level variables. Finally, we estimate the impact of UI over a different time period than HKMM that additionally includes the phaseout of benefit extensions. Implementing each of these three changes substantially reduces the estimated magnitude of the negative effect on employment, and as a result, many of our point estimates are in fact positive.[4] We also discuss HMM in detail. We show how the results in HMM are not robust to data revisions issued by the Bureau of Labor Statistics and rely upon particular parametric assumptions about counterfactual employment trends.

There have been other criticisms of the BCP approach, which we show to be quantitatively less important. In particular, Hall (2013) raises concerns that certain

[3] HKMM find that the expansion of UI during the Great Recession from 26 to 99 weeks increased the unemployment rate by 80 percent, which is an effect on unemployment that is roughly comparable to the unemployment growth that actually occurred during the Great Recession itself; they interpret this result as an explanation for the slow recovery in the unemployment rate in the years after the trough of the Great Recession. HMM study the 2014 expiration of EUC and find that the expiration was responsible for the creation of approximately two million jobs. This effect would translate into a 1.1 percent decrease in employment as a result of the expansion of UI from 26 to 99 weeks, which corresponds to about one-third of the employment decrease of the Great Recession as measured in our dataset.

[4] Amaral and Ice (2014) also demonstrate the sensitivity of HKMM's results to the sample period; they show how the use of a longer sample period leads to a much smaller estimate than HKMM. We show that this choice of sample period is one of multiple reasons why the HKMM estimate is overstated. We also explicitly examine the robustness of our estimates to the choice of sample period.

counties have employment that is highly correlated (potentially mechanically) with state-level outcomes, which would undo the benefit of the BCP strategy. Yet, we find our estimates unchanged when we drop counties whose employment is highly correlated with state employment. We also find no significant differences in county-level covariates within border county pairs, nor do we find pretrends in the form of statistically significant leads in dynamic regressions.

Dieterle, Bartalotti, and Brummet (2020) also raise concerns about using neighboring counties as controls, given that counties are large and much of the population within a county can live far away from the border. They develop an innovative regression discontinuity approach that takes into account the spatial distribution of the population within a county and controls for distance to the state boundary. Applying this modified border county approach, they find that the HKMM estimates are substantially reduced. However, one limitation is that the standard errors from the Dieterle, Bartalotti, and Brummet (2020) specification are wide enough to include the estimate of HKMM. Importantly, we show that the addition of county fixed effects to the Dieterle, Bartalotti, and Brummet (2020) specification (important when doing a difference-in-discontinuity-style estimation) both substantially reduces the standard errors and stabilizes the point estimates to the addition of spatial controls. When county fixed effects are included, the RD and the BCP estimates for employment are essentially the same—with larger standard errors for the RD specification. We thus see the spatial RD approach and the BCP approach as providing complementary evidence.

There are several other recent papers that also have estimated a macro impact of UI on employment. As discussed above, Chodorow-Reich, Coglianese, and Karabarbounis (2019) find small and statistically insignificant effects, similar to ours, of UI extensions on aggregate employment using a methodology that exploits real-time measurement error in state unemployment rates. Since UI duration was tied to the state unemployment rate, two states with similar labor market conditions could have different UI duration because of errors in the measured unemployment rate. One limitation of this measurement-error-based approach is that the policy changes they study are less durable than the changes we examine in this paper, and thus the external validity may be more limited. Nonetheless, the very different types of variation leveraged across our two sets of papers make them complementary. Our findings are also consistent with Marinescu (2017), who finds that UI benefit extensions during the Great Recession decreased job applications but not posted vacancies, implying a modest impact of the extensions on overall job finding and unemployment rates. Finally, in their case study of Missouri, Johnston and Mas (2018) employ a synthetic control estimator and find substantially larger negative macro employment effects than we find in this paper. However, their micro estimates are also substantial. Ultimately, they find a gap between their micro and their macro estimates that is comparable to the gaps we find relative to most of the micro estimates from the Great Recession in the United States. Our approach differs from their macro estimates primarily in that we aggregate across many different benefit extensions and reductions and that our analysis uses variation across border counties rather than neighboring or similar states.

The remainder of the paper is structured as follows. In Section I, we discuss important institutional details of the UI extensions during the Great Recession that are critical for our identification strategy. In Section II, we discuss our data. In Section III, we discuss our identification challenges and provide evidence in support of the BCP approach. In Section IV, we present our empirical results. Section V compares our findings to those in HKMM and HMM. In Section VI, we compare our macro estimates of UI expansion on employment with micro-level estimates based on labor supply elasticities. Finally, in Section VII, we conclude.

I. Unemployment Insurance Background

During the Great Recession, the maximum potential duration of UI receipt increased from 26 weeks (in all but two states) to as much as 99 weeks.[5] The path of this expansion and subsequent phaseout differed across states, which creates the variation that we use in this study.

Three sets of policies contribute to the variation in maximum benefit duration across states. The most quantitatively important policy generating cross-state variation was the Emergency Unemployment Compensation (EUC) program, signed into law on June 30, 2008. Initially, EUC increased the maximum UI duration by 13 weeks in all states. This program was modified several times. Beginning in November 2008 and continuing until its expiration at the end of 2013, EUC provided more weeks of benefits in states with a higher unemployment rate. For example, between November 2009 and February 2012, 34 weeks of EUC benefits were generally available in all states, while states with high unemployment rates were eligible for those 34 weeks plus an additional 13 or 19 weeks, depending upon the unemployment rate. A bill passed in February 2012 began reducing benefit duration in May of that year, and the EUC program phased out fully on January 1, 2014. Differential EUC benefits provide the majority of the across-state variation over our sample period.

Additionally, variation is created by a preexisting policy known as Extended Benefits (EB). EB, available continuously since passage in 1970, provides 13 or 20 weeks of additional UI benefits in a given state if that state's unemployment rate is high enough (or growing fast enough) to exceed certain "triggers." Notably, some of these triggers were optional—meaning that state policymakers could choose whether or not to adopt them. Despite the fact that during the Great Recession, the federal government paid for the entirety of EB, many states chose not to adopt any of the optional triggers and did not become eligible for EB at any point during the Great Recession. Thus, variability in EB across states is driven both by changes in state-specific unemployment rates and persistent differences in state policy interacted with national changes in unemployment rates. We provide greater details about the EB and EUC programs in online Appendix A.

[5] Before the Great Recession, Massachusetts and Montana both had more than 26 weeks of UI; Montana had a maximum of 28, and Massachusetts had a maximum of 30. In contrast to the UI expansions during COVID-19, there were no large federally driven increases in weekly benefit amounts (beyond a temporary \$25 per week increase contained in the 2009 Recovery Act).

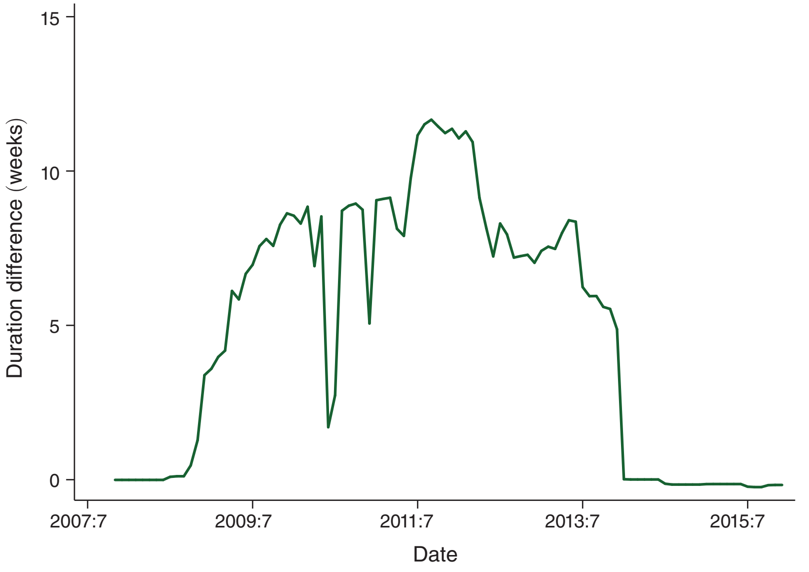

Figure 1. Difference in UI Benefit Duration between High-Treatment and Low-Treatment Counties across State Borders

Notes: For each county pair, we compute the difference between maximum duration in the high-duration county and the low-duration county. We plot the average difference across all county pairs. "High" and "low" status is determined by comparing the difference between average duration from 2008:11–2013:12 and average duration from 2007:11–2008:10 and 2014:1–2014:12. The counties in the 30 pairs where this difference is identical are assigned arbitrarily to the "high" and "low" sets.

Finally, between 2011 and the end of our sample period, several states reduced their maximum duration of regular benefits below the usual 26 weeks, contributing to across-state variation. Most notably, North Carolina reduced their regular weeks of benefits in the middle of 2013; at the same time, other changes to North Carolina's UI system caused the EUC in North Carolina to immediately lapse. All told, the maximum benefit duration in North Carolina fell immediately by 53 weeks.[6] While such state-level reforms are not mechanically caused by changes in economic condition, it is likely that economic conditions (especially the state of the state UI trust fund) played a role in these policy decisions.

The differences in UI benefits across state lines were substantial. In Figure 1, we show the differences across neighboring counties in the numbers of weeks of available UI, where the reported difference is between "high" and "low" benefit duration counties, defined by comparing the average duration in the treatment period (2008:11–2013:12) versus the surrounding 24 months (2007:11–2008:10 and 2014:1–2014:12) when these differences were typically 0 or very small.

[6] In addition, in 2011, Arkansas reduced its maximum benefit duration to 25 weeks, and both Missouri and South Carolina reduced theirs to 20 weeks. Then, in 2012, Florida, Georgia, Illinois, and Michigan reduced their maximum benefit duration: Michigan lowered it to 20 weeks, while the other three made it contingent on the state unemployment rate.

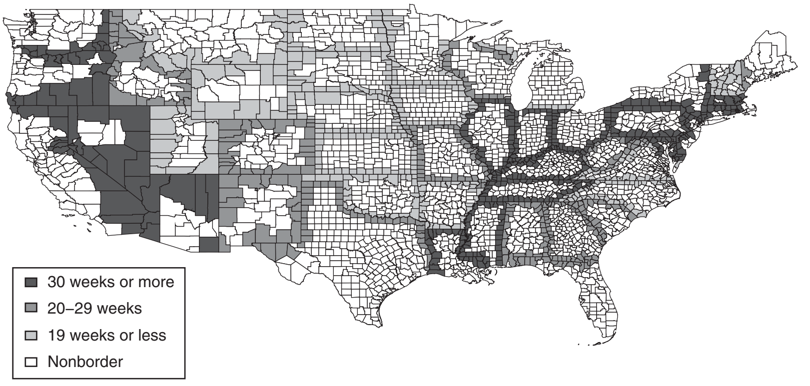

Figure 2. Reduction in UI Benefit Duration from the December 2013 Expiration of EUC

The average gap between states with longer versus shorter total duration within the county pairs rose to nearly 12 weeks by late 2011 before declining to an average gap of near 0 with the expiration of EUC in December 2013. This variation over time is used in our full panel estimates. We also use the national-level policy variation due to the November 2008 expansion and the late-2013 expiration of the EUC program as instruments for our event study strategy. In Figure 2, we show a map of the counties that had different generosity levels right before the EUC expiration in December 2013. Online Appendix Figure C1 shows the analogous map for the variation created by expansion of the EUC program in November 2008.

II. Data

We use county-level employment data from the Quarterly Census of Employment and Wages (QCEW, Bureau of Labor Statistics 1998–2015). The QCEW data is based on ES-202 filings that nearly all establishments are required to file on a quarterly basis with their state government for the purpose of calculating UI-related payroll taxes. These employment and earnings counts are shared by the states with the Bureau of Labor Statistics, which releases the data at the county-industry-quarter level. Since 98 percent of jobs are covered by UI, these payroll counts constitute a near census of employment and earnings.[7] The QCEW provides total employment for each month at the county level. In our baseline estimation, we require that each county be in the dataset in every month. This excludes four counties for which there is at least one month in the sample where the QCEW does not report data due to confidentiality problems with disclosure. This occurs only in counties with very low population. In our robustness section, we additionally report estimates using the full unbalanced panel.

[7] The QCEW covers both private and public sector employment. We focus our analysis on total employment (the sum of private and public sector employment), though we do provide results on private employment as a robustness check. There are some limitations to the data: the QCEW does not capture workers in the informal sector or the self-employed, and it misses the small number of workers who participate in their own UI system, such as railroad workers and workers at religiously affiliated schools.

We divide employment by population of those 15 and older, which we obtain from the census at the annual level and interpolate log-linearly within each year (US Census Bureau 2016a). Prior to estimation, we seasonally adjust our dependent variables by subtracting off the county-month specific mean of the variable in question, where this mean is calculated out of sample over the period 1998–2004.[8] As we show later in the paper, however, our results are robust to using raw rather than seasonally adjusted data.[9]

Our data on the number of weeks of regular benefits comes from Department of Labor reports that are issued biannually (Employment and Training Administration 1990–2015b). To account for occasional changes in the numbers of weeks of regular benefits that occur during the intervening period, we augment these data with online searches of news media and state government websites. We obtain information on EUC and EB from the trigger reports released by the Department of Labor (Employment and Training Administration 2003–2015a, 2008–2013). These reports provide the number of weeks of EB and tiers of EUC available for each state in each week. When a change in weeks of benefits happens within a month, we assign the time-weighted average of the maximum duration to that month.

As discussed above, there were several lapses in the EUC program during 2010. In the popular press, expectations were that these lapses would be reversed and the original EUC benefit durations would be reinstated. In each of these cases, this is in fact what did happen. In our baseline specifications, we treat the lapses as true expirations—that is, those county-by-month observations are coded as having EUC equal to zero. We show in robustness checks that our estimates are not substantially affected if we code the benefit durations for these few months as having remained unchanged at their prelapse level.

We also use a list of all contiguous county pairs that straddle state borders; these data come from Dube, Lester, and Reich (2010). In our baseline specifications, we have a total of 1,161 county pairs.

In addition, we obtain county-level unemployment and employment data at the quarterly level from the Local Area Unemployment Statistics (LAUS) published by the Bureau of Labor Statistics. We obtained the most current data (as of November 10, 2016) via http://download.bls.gov/pub/time.series/la/ (Bureau of Labor Statistics 1990–2015). We additionally obtain a vintage series of county unemployment rates and employment (prior to the March 2015 redesign) via FRED (Bureau of Labor Statistics 1976–2015). The pre-redesign LAUS data is the source for the unemployment rate variable used by HKMM and HMM, and we use it as part of our reconciliation exercise in Section V.

[8] For the sake of summary statistics and the small number of specifications we estimate without county fixed effects, we add back the overall mean level of EPOP for each county measured over the 1998–2004 period.

[9] The time series of aggregate EPOP is displayed in online Appendix Figure C2 together with the average maximum UI duration over time. Our measure of EPOP is smaller than the US DOL measure. This is largely because our measure is based upon UI employment and thus excludes those in the informal sector as well as the self-employed. Additionally, we calculate EPOP by dividing employment by the 15+ population in the county rather than the 16+ population used by the DOL.

III. Border-County-Pair Design and Graphical Evidence

In order to credibly estimate the effect of UI extensions on aggregate employment, we need to address a severe problem of reverse causality. Because UI benefit levels were tied to state unemployment rates, negative employment shocks that increased unemployment were likely to mechanically raise the maximum benefit duration as well. The presence of this "mechanical endogeneity" motivates us to restrict our sample to contiguous county pairs that straddle state borders (Dube, Lester, and Reich 2010, 2016) and estimate the effects within border county pairs. The main idea behind this strategy is that neighboring counties in adjacent states are reasonably well matched and likely to face similar economic shocks, yet their UI maximum benefit durations will be driven by their respective states' unemployment rates and policy choices—which may be quite different. Therefore, by focusing on comparisons between border counties, we are able to account for confounders that vary smoothly across state borders and better account for the mechanical endogeneity problem that would plague a two-way (state and time) fixed effects approach.

We begin by showing pretreatment covariate balance within border county pairs and the absence of preexisting trends within border pairs at an aggregate level. In particular, for each month t, we organize our border-county-pair (BCP) data to have two observations in each pair p—one for each county c of the pair. This means that a given county c appears in the data k times (for each month t) if it borders k counties in adjacent states. We then define a county-specific time invariant measure of treatment, $treat_c$, which we define as the difference in time-averaged maximum benefit duration in a given county during the "treatment period" (i.e., between November 2008 and December 2013) versus the 12 months prior (i.e., between November 2007 and October 2008) and 12 months after (i.e., between January and December 2014).[10] For example, if a state's average maximum UI duration during the treatment period was 90 weeks and the average maximum benefit length in the 12 nontreatment months was 30 weeks, then it would have a value of $treat_c$ equal to 60 weeks. For ease of interpretation, we rescale this variable by dividing it by 10 so that a value of 1 corresponds to a difference of 10 weeks of treatment, which is roughly equal to the mean difference in duration between neighboring counties that straddle state borders during the treatment period. Table 1 shows that the "high" and "low" treatment counties were quite similar. Preexisting characteristics are relatively balanced within pairs between the counties that received high versus low average treatments. Only one of 12 covariates (share with a college degree) is statistically significantly different at conventional levels.[11]

[10] This "nontreatment" value will in general not be equal to 26—the usual maximum duration for regular benefits—since it includes the period from July to October 2008 when all states were eligible for 13 weeks of EUC.

[11] In online Appendix Table C1, we compare differences in the border pair sample to differences in pairs formed by randomly matching counties. Differences are substantially smaller in the border sample, especially with respect to population, urban share, and amount of mortgage debt originated in 2007. Summary statistics in online Appendix Table C2 also show that border counties are relatively comparable to the full set of counties, indicating that the sample restriction for purposes of internal validity comes at minimal sacrifice of external validity.

Table 1—Summary Statistics: High-Treatment versus Low-Treatment Counties in Border-County-Pair Sample

	Baseline			PT-Trimmed		
	High	Low	p-value	High	Low	p-value
EPOP (A)	44.107	44.621		42.490	43.730	
	(17.044)	(15.110)	0.668	(14.545)	(13.422)	0.185
Private EPOP (A)	32.742	33.485		31.590	32.908	
	(14.468)	(14.074)	0.476	(12.653)	(12.500)	0.106
LAUS unemployment rate (A)	7.745	7.167		7.924	7.310	
	(2.358)	(2.403)	0.006	(2.182)	(2.233)	0.000
Population age 15+ (A)	81,415	71,441		96,552	80,856	
	(213,875)	(153,060)	0.211	(240,297)	(155,185)	0.015
Share White (B)	0.811	0.811		0.814	0.813	
	(0.182)	(0.177)	0.998	(0.179)	(0.176)	0.946
Share Black (B)	0.085	0.086		0.090	0.088	
	(0.145)	(0.147)	0.966	(0.147)	(0.144)	0.834
Share Hispanic (B)	0.067	0.059		0.060	0.053	
	(0.111)	(0.092)	0.491	(0.100)	(0.086)	0.468
Share high school grad (B)	0.569	0.567		0.567	0.565	
	(0.064)	(0.065)	0.724	(0.062)	(0.065)	0.691
Share college (B)	0.179	0.189		0.182	0.191	
	(0.078)	(0.086)	0.010	(0.081)	(0.087)	0.001
Median household income (B)	42,645	43,535		42,898	43,969	
	(11,459)	(12,127)	0.198	(11,937)	(12,775)	0.119
New mortgage debt per capita (A)	3.386	3.586		3.488	3.687	
	(3.092)	(2.877)	0.447	(2.874)	(2.924)	0.435
Share in cities 50k+ (C)	0.190	0.196		0.214	0.230	
	(0.331)	(0.331)	0.759	(0.345)	(0.348)	0.396
Minimum weeks of UI eligibility	24.470	24.631		24.436	24.787	
	(3.495)	(3.199)	0.718	(3.511)	(3.030)	0.494
Maximum weeks of UI eligibility	96.105	86.996		96.277	87.659	
	(6.674)	(13.320)	0.000	(6.560)	(12.622)	0.000
Pairs with different average treatment	1,131	1,131		848	848	
Pairs with identical average treatment	30	30		22	22	

Notes: The first two columns report means and (in parentheses) standard deviations for border counties in the estimation sample, separately for "high" and "low" treatment counties. A county's assignment to the "high" or "low" group is defined by its average treatment intensity relative to its counterpart within each pair. Average treatment intensity ($treat_c$) is a time-invariant, continuous measure defined as the average duration over the 2008:11–2013:12 period minus average duration over the 2007:11–2008:10 and 2014:1–2014:12 periods. The 30 (baseline) or 20 (PT-trimmed) border county pairs with identical treatment are dropped in this table. The third column reports the p-values from a test that the means for high counties and low counties are equal, robust to clustering two-way at the state and state-pair level. Columns 4–6 report analogous statistics for the subsample of border county pairs in the PT-trimmed sample. PT-trimming removes the quartile of county pairs with the highest mean squared error in EPOP between November 2004 and October 2008 (after partialling out a fixed level difference). If a border county appears in j county pairs, then it appears j times for the purpose of creating the estimates in this table. (A) is from 2007 data, (B) is from the 2005–2009 ACS, and (C) is from the 2010 census. High school graduates are those who have attained a high school degree but not a bachelor's degree. College graduates are those who have attained a bachelor's degree.

We also present aggregate graphical evidence on preexisting trends and treatment effects. In particular, we regress EPOP on a set of $treat_c \times \mathbf{1}\{t = s\}$ variables, where $\mathbf{1}\{t = s\}$ is an indicator for date s, controlling for county fixed effects λ_c. In the full sample, we normalize estimates to October 2008 by omitting the variable corresponding to that month. We additionally control for county fixed effects λ_c and pair-period effects ν_{pt}. The estimating equation is as follows:[12]

$$(1) \qquad E_{cpt} = \sum_{s=\tau_A}^{\tau_B} \beta_s \, treat_c \mathbf{1}\{t = s\} + \lambda_c + \nu_{pt} + \epsilon_{cpt}.$$

Since $treat_c$ is a continuous, time-invariant measure, the coefficients β_s trace out how EPOP evolves in the treated versus control sides over time as compared to a base period of October 2008, the month before the first cross-state variation in federal UI benefits in our sample. The pair-period effects, ν_{pt}, sweep out the variation between pairs, leaving only the within-pair variation to identify β_s.[13]

To illustrate the role of local comparisons, we first present a figure where we replace the pair-specific period fixed effects with common period fixed effects. This classic two-way fixed effects specification compares EPOP in counties with higher average duration to others with lower average duration. Figure 3 plots the coefficient estimates.[14] The figure plots two sets of coefficients: one with EPOP as the dependent variable in blue and the other with maximum UI benefit duration as the dependent variable in red. There is clear evidence of the mechanical endogeneity: treated counties averaging 10-week-higher UI benefit duration during the Great Recession experienced slightly more than 1 percentage point relative decline in EPOP in the 3.5 years prior to November 2008.

Figure 4 plots analogous regression coefficients but now with pair-period fixed effects, therefore only using variation *within neighboring county pairs.* The main result of our paper is visible in this figure. First, in contrast to the classic two-way fixed effects specification, there is no statistically significant (or sizable) preexisting trend once we compare within border county pairs.[15] There is some relative decline in EPOP starting in 2009. Had we only focused on the sample through the middle of 2011, the rise in benefit duration and fall in EPOP would suggest some negative effect of UI benefits, though the estimates are not distinguishable from zero.

[12] Except where noted, our standard errors are clustered two-way at the state-pair level and at the state level. Clustering at the state-pair level is designed to account for common, serially correlated shocks to local economies. We also cluster at the state level to account for the mechanical correlation in error terms that is introduced when one county borders counties in at least two states (and thus appears in multiple state pairs), as well as any state-level shocks. Note that our clustering strategy fully accounts for the appearance of a single state multiple times in the border-county-pair sample.

[13] With two observations within each pair-period group, this approach gives the identical coefficients as if we dropped the pair-period fixed effects and instead (i) took the spatial difference of the dependent variable and main independent variable across each county pair p at each time t and (ii) replaced county fixed effects with pair fixed effects.

[14] Online Appendix Figure C3 plots an analogous figure using data from all counties, which is qualitatively similar. Thus, in the absence of the pair-period fixed effects, the organization of the data into border county pairs is largely immaterial.

[15] Online Appendix Table C3 quantifies the magnitude of these preexisting trends by regressing EPOP on $treat_c \times t$ in the four years prior to treatment, conditional on county fixed effects and period fixed effects (in columns 1 and 2) and pair-period fixed effects (in columns 3 and 4). This table shows that the preexisting trends are negative and significant in specifications without pair-period fixed effects and less negative (and insignificant) in column 3, the specification analogous to Figure 4. Furthermore, this evidence is complementary with the evidence provided in Section 4.3 of HKMM. HKMM find substantially larger estimates of the effect of UI on unemployment when their border pair sample is replaced by a "scrambled border pair" sample, in which pairs are formed randomly (rather than by reason of geographical adjacency). HKMM argue (and we agree) that this is indicative of the role played by the BCP strategy in reducing mechanical endogeneity.

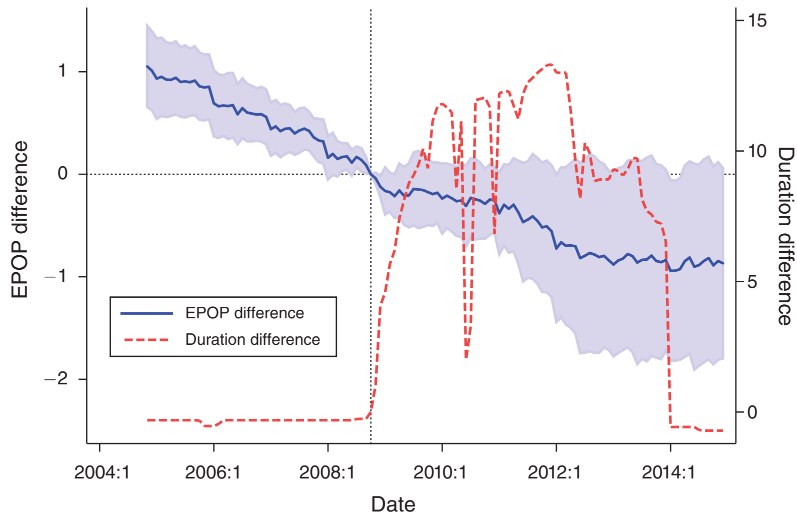

FIGURE 3. EVOLUTION OF EPOP AND UI BENEFIT DURATION DIFFERENTIALS BY AVERAGE TREATMENT INTENSITY: WITHOUT PAIR-PERIOD FIXED EFFECTS

Notes: This figure plots (solid line, left axis) the set of β_s coefficients from the following regression: $E_{cpt} = \sum_{s=\tau_A}^{\tau_B} \beta_s treat_c \mathbf{1}\{t = s\} + \lambda_c + \nu_t + \epsilon_{cpt}$. E_{cpt} is the seasonally adjusted ratio of total employment to population age 15+, scaled in percentage points. The average treatment intensity, $treat_c$, is a time-invariant, continuous measure defined as the average duration during the treatment period (2008:11–2013:12) minus average duration from the 12 months prior (2007:11–2008:10), divided by 10. The shaded region corresponds to the 95 percent confidence interval, robust to two-way clustering at the state and state-pair level. The dotted line (right axis) reflects the analogous coefficients with D_{ct} as the dependent variable, where D_{ct} is weeks of benefits. The month 2008:10, the last month prior to the first introduction of differential EUC, is marked with a dotted vertical line. The sample includes 1,161 county pairs.

Contrary to that interpretation, employment continued to fall in the treated side past 2011 through 2014, a period when the treatment difference declines due to rollback of federal benefits, leaving little correlation between benefit duration and employment in the full sample. This previews our regression results that overall employment effects of UI benefits are likely to be modest.

Trimming on Pretreatment Match Quality.—While the evidence in Figure 4 shows that the BCP design mostly eliminates the preexisting trends that afflict the two-way fixed effects model, we additionally consider a refinement where we drop pairs with the largest within-pair variation in presample EPOP. Restricting the analysis to pairs that are better matched based on pretreatment characteristics may help further reduce the influence of unobserved heterogeneity. We consider three different match quality criteria and then use a data-driven approach to select between these three. These criteria are as follows: (i) pretreatment mean squared error (MSE) in employment between counties, (ii) absolute differences in pretreatment linear time trend in employment between counties, and (iii) differences in pretreatment demographic and economic covariates between counties.[16] For each of these three criteria, we

[16] In particular, we construct a Mahalanobis measure incorporating within-pair differences in share white, share Hispanic, share with a bachelor's degree or more, log median per capita income, share urban, number of mortgage

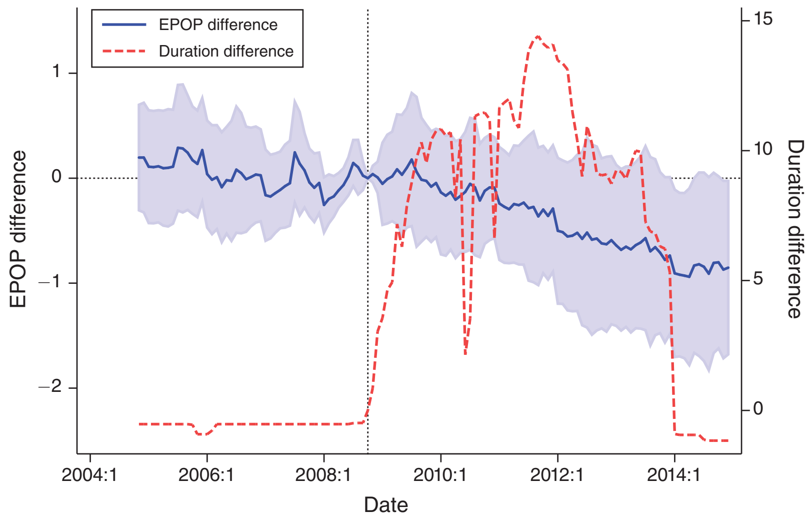

FIGURE 4. EVOLUTION OF EPOP AND UI BENEFIT DURATION DIFFERENTIALS BY AVERAGE TREATMENT INTENSITY: BASELINE BORDER-COUNTY-PAIR SAMPLE

Notes: This figure plots (solid line, left axis) the set of β_s coefficients from the following regression: $E_{cpt} = \sum_{s=\tau_A}^{\tau_B} \beta_s treat_c \mathbf{1}\{t = s\} + \lambda_c + \nu_{pt} + \epsilon_{cpt}$. E_{cpt} is the seasonally adjusted ratio of total employment to population age 15+, scaled in percentage points. The average treatment intensity, $treat_c$, is a time-invariant, continuous measure defined as the average duration during the treatment period (2008:11–2013:12) minus average duration from the 12 months prior (2007:11–2008:10), divided by 10. The shaded region corresponds to the 95 percent confidence interval, robust to two-way clustering at the state and state-pair level. The dotted line (right axis) reflects the analogous coefficients with D_{ct} as the dependent variable, where D_{ct} is weeks of benefits. The month 2008:10, the last month prior to the first introduction of differential EUC, is marked with a dotted vertical line. The sample includes 1,161 county pairs.

estimate a "match quality" variable in the "training sample" of 2003:11–2007:10. Next, we trim the worst 25 percent of the pairs based on each criteria. Finally, we use a "test sample" of the 2007:11–2008:10 period prior to treatment to assess the out-of-sample MSE in employment. We choose the trimming criteria that have the minimal out-of-sample MSE. We find that pretreatment MSE trimming performs the best: as compared to the baseline BCP sample, it reduces out-of-sample MSE by 65 percent. In contrast, trimming on pretreatment linear trends reduces out-of-sample mean MSE by 44 percent, and covariate trimming results in a 25 percent reduction. Therefore, throughout the paper, we show estimates based on the baseline BCP-FE sample as well as the pretreatment-trimmed (henceforth PT-trimmed) sample that drops the county pairs in the top quartile of pretreatment MSE.

Figure 5 shows results analogous to Figure 4 using our refined PT-trimmed set of border county pairs, where we exclude the pairs with the largest differences in preexisting MSE. The confidence intervals are substantially narrower in the latter figure. In addition, we note that the relative downward trend in EPOP for treated counties is muted in the PT-trimmed sample: employment differences are quite

originations in 2007, log population, 2006 EPOP, and share of employment in the goods industry.

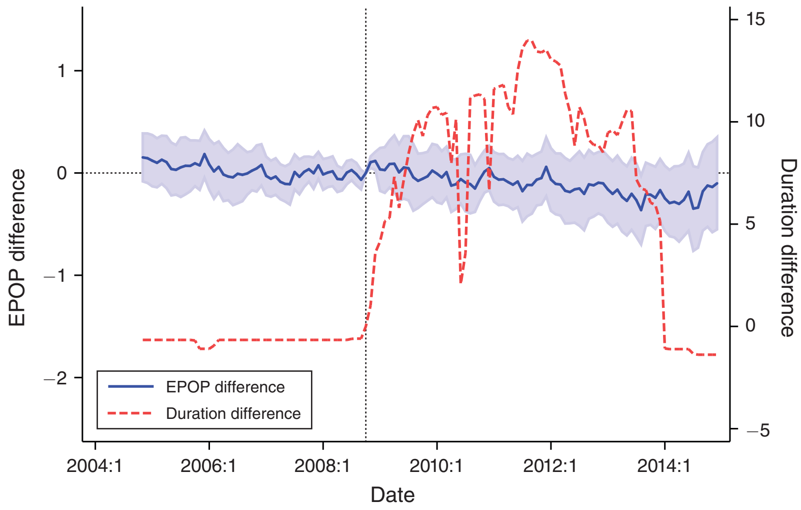

Figure 5. Evolution of EPOP and UI Benefit Duration Differentials by Average Treatment Intensity: PT-Trimmed Border-County-Pair Sample

Notes: This figure plots (solid line, left axis) the set of β_s coefficients from the following regression estimated over the set of border county pairs in the PT-trimmed sample: $E_{cpt} = \sum_{s=\tau_A}^{\tau_B} \beta_s treat_c \mathbf{1}\{t = s\} + \lambda_c + \nu_{pt} + \epsilon_{cpt}$. E_{cpt} is the seasonally adjusted ratio of total employment to population age 15+, scaled in percentage points. The average treatment intensity, $treat_c$, is a time-invariant, continuous measure defined as the average duration during the treatment period (2008:11–2013:12) minus average duration from the 12 months prior (2007:11–2008:10), divided by 10. The shaded region corresponds to the 95 percent confidence interval, robust to two-way clustering at the state and state-pair level. The dotted line (right axis) reflects the analogous coefficients with D_{ct} as the dependent variable, where D_{ct} is weeks of benefits. The month 2008:10, the last month prior to the first introduction of differential EUC, is marked with a dotted vertical line. PT-trimming removes the quartile of county pairs with the highest mean squared error in EPOP between November 2004 and October 2008 (after partialling out a fixed level difference). The sample includes 870 county pairs.

stable during the entire 2005–2014 period. At the same time, the estimates based on the BCP-FE and the PT-trimmed samples are unlikely to differ much—as discussed above, the downward employment trend in Figure 4 occurs during both benefit expansion and contractions, implying employment and UI benefits are largely uncorrelated for both the baseline BCP-FE and PT-trimmed samples.

IV. Empirical Findings

While the time-invariant aggregate treatment measure is useful for a qualitative, visual assessment of how employment evolved on the two sides of the border, it does not make use of the timing of changes in UI generosity across states. This section presents the results from our main empirical specifications, which compare changes in UI benefit duration to changes in county-level employment within county pairs over time. We first present the results from estimating our border-county-pair fixed effects (BCP-FE) specification over the entire sample period and making use of all cross-border variation in UI benefit duration. Then we present the results from our "event study" strategy that exploits only the variation induced by national-level policy changes to the EUC program in 2008 and 2014.

A. *Full Sample Results*

Our baseline BCP-FE specification equation uses a normalized maximum benefit duration, D_{ct}, to estimate the following equation:

$$E_{cpt} = \beta D_{ct} + \lambda_c + \nu_{pt} + \eta_{cpt}. \tag{2}$$

We normalize D_{ct} by dividing the maximum benefit duration (in weeks) by 73 to make β interpretable as the change in EPOP resulting from an increase equal to the maximum expansion that occurred during the Great Recession.[17] We include pair-period effects ν_{pt} to sweep out between-pair variation and county fixed effects λ_c to account for persistent differences between the two members of the pair.[18] This strategy relies on D_{ct} being uncorrelated with η_{cpt}, i.e., $E(D_{ct}\eta_{cpt}) = 0$, but this assumption needs to hold only within a local area that is likely to be experiencing more similar economic shocks. Figures 3 and 4, respectively, show that this assumption is not likely to hold unconditional upon pair-period fixed effects but is more likely to hold conditional upon pair-time effects.

We present our full sample estimates in the panel A of Table 2. This panel reports two columns of regressions estimating equation (2). The first column reports results using the baseline (i.e., untrimmed) BCP sample, and the second column reports results using the sample that we refined based on pretreatment MSE (the PT-trimmed sample). The point estimate for the baseline BCP sample is 0.430. Given the scaling of D_{ct}, this coefficient estimate represents the estimated impact on EPOP from an increase in maximum benefit duration from 26 to 99 weeks. Consequently, the baseline BCP estimate suggests that the 73-week increase in maximum benefit duration raised the EPOP ratio by 0.430 percentage points. The standard error is 0.471, and thus the estimate is far from statistically distinguishable from zero. When we restrict the analysis to the PT-trimmed sample in column 2, the coefficient falls to 0.180, with a smaller standard error of 0.268. Because of the smaller standard error in the PT-trimmed sample, the bottom of the 95 percent confidence interval falls in magnitude from −0.493 to −0.345. The upper limits of the confidence interval change from 1.353 with the full sample to 0.705 in the trimmed sample.

Dynamic Evidence.—We also present dynamic estimates of the employment effect of UI duration around the time of policy change. There are two specific aims that underlie this analysis. First, we wish to use the leading coefficients to detect preexisting trends and assess the validity of the research design. Second, we wish to assess possible anticipation or lagged effects of the policy. In particular, we will interpret small, statistically insignificant lead coefficients and an absence of trend in the lead coefficients as evidence against market anticipation of future changes in UI generosity. To this end, we utilize a first-differenced distributed lag specification

[17] Thirty-two of the 51 states (including DC) experienced an increase in duration of 73 weeks.

[18] Formally, the ν_{pt} are actually county-cross-county-pair fixed effects. This distinction mechanically makes no difference except in the small number of specifications in which the panel is unbalanced. In these cases, it subtracts county means from a county only for time periods when the county pair is in the dataset and thus potentially differently in some pairs compared to others.

Table 2—Main Estimates: Effect of UI Benefit Duration on EPOP Using Full Sample and Event Study Specifications

	BCP-FE (1)	PT-Trimmed (2)
Panel A. Full sample		
OLS estimate	0.430	0.180
	(0.471)	(0.268)
County pairs	1,161	870
Observations	199,692	149,640
Panel B. Pooled sample (IV)		
IV estimate	0.143	0.253
	(0.974)	(0.650)
First-stage coefficient	0.847	0.842
	(0.052)	(0.052)
F-statistic	[262.2]	[262.3]
County pairs	1,161	870
Observations	108,000	80,928
Panel C. 2008 sample (IV)		
IV estimate	0.549	1.344
	(2.541)	(1.253)
First-stage coefficient	0.717	0.726
	(0.111)	(0.114)
F-statistic	[41.3]	[40.3]
County pairs	1,161	870
Observations	55,728	41,760
Panel D. 2014 sample (IV)		
IV estimate	−0.024	−0.214
	(0.568)	(0.523)
First-stage coefficient	0.915	0.903
	(0.046)	(0.044)
F-statistic	[392.6]	[423.8]
County pairs	1,089	816
Observations	52,272	39,168

Notes: Each panel reports two coefficients on D_{ct} from a regression of the form $E_{cpt} = \beta D_{ct} + \lambda_c + \nu_{pt} + \eta_{cpt}$. E_{cpt} is the seasonally adjusted ratio of total employment to population age 15+, scaled in percentage points, and D_{ct} is the potential weeks of UI benefits divided by 73. The second column restricts the sample to the PT sample. PT-trimming removes the quartile of county pairs with the highest mean squared error in EPOP between November 2004 and October 2008 (after partialling out a fixed level difference). Regressions in panel A use OLS estimated over the 2007:11–2014:12 period. Regressions in the remainder of the table are estimated on subsamples using instrumental variables. The instrument z_{ct} is defined as follows. From 2007:11–2008:10, z_{ct} is equal to the duration available immediately prior to the implementation of UCEA; from 2008:11–2009:10, z_{ct} is equal to the duration available immediately after the implementation of UCEA. From 2013:1–2013:12, z_{ct} is equal to the duration available immediately prior to the expiration of EUC; from 2014:1–2014:12, z_{ct} is equal to the duration available immediately after EUC expiration, before any changes in regular benefits took effect. Estimates in panel B pool the 2007:11–2009:10 and 2013:1–2014:12 samples and replace county fixed effects with county-by-subsample fixed effects. Estimates in panel C use data from 2007:11–2009:10; estimates in the panel D use data from 2013:1–2014:12. In the IV specifications, first-stage coefficients and standard errors are also reported. Standard errors are reported in parentheses and first-stage *F*-statistics are in square brackets. Standard errors are two-way clustered at the state and state-pair level.

with a set of 11 monthly leads and 24 monthly lags, along with the contemporaneous benefit duration, D_{ct}. This specification allows us to focus on employment changes within the 36-month window around the time of treatment.

Our estimating equation for the dynamic specification is

$$\Delta E_{ct} = \sum_{k=-11}^{24} \beta_k \Delta D_{c,t-k} + \nu_{pt} + \epsilon_{cpt}. \tag{3}$$

Successively summing the coefficients traces out the cumulative response to a one-time, permanent unit change in D: $\rho_\tau = \sum_{k=-11}^{\tau} \beta_k$ represents the cumulative response at event time, τ.[19] For ease of interpretation, we center the cumulative responses around a baseline of the month just prior to treatment, $\tilde{\rho}_\tau = \rho_\tau - \rho_{-1}$, which imposes that $\tilde{\rho}_{-1} = 0$. We plot the centered cumulative response $\tilde{\rho}_\tau$ by event time along with the associated confidence intervals below.

Figure 6 visually displays the results of the first-differenced distributed lag specification of equation (3).[20] The figure shows the cumulative response in employment $(\tilde{\rho}_\tau)$ starting 12 months before treatment and extending up to 24 months after, relative to the month before treatment. The top panel displays the coefficients for the full sample of BCPs, while the bottom panel displays them for the PT-trimmed sample. For both specifications, during the 12 months prior to treatment (i.e., between $\tau = -12$ and -1), there is no statistically or economically significant change in employment. The twelfth lead is $+0.228$ in the full sample and a mere $+0.059$ in the trimmed sample. The leading values of the cumulative responses range between -0.083 and $+0.403$ and are never statistically distinguishable from zero. Overall, the distributed lag specifications produce little evidence to indicate reduced hiring in anticipation of the policy change.

Following treatment, both the baseline BCP specification and the PT-trimmed specification show no change in employment over the 24 months following the policy change. The cumulative responses are typically positive and not statistically significantly different from zero. Even as the precision declines for longer lags, 12 months after the policy change, we can nonetheless still rule out employment effects more negative than -0.6 with 95 percent confidence for both specifications. Overall, the dynamic evidence suggests little employment change in the year prior to treatment (e.g., through anticipation) or during the two years following the policy change.

B. *Estimates from the EUC Expansion and Expiration Events*

Estimating equation (2) over the full sample period exploits all of the variation in maximum benefit duration induced by both policy changes[21] as well as movements in state unemployment rates across various thresholds. This latter source of UI variation, resulting from the presence of unemployment rate "triggers" in the design of the EB and EUC policies, may introduce endogeneity bias even within neighboring county pairs that lie in different states. Our use of the BCP design helps

[19] Note that β_k is the response associated with D_{t-k}. This indexation convention allows us to index the coefficients by event time.

[20] The exact values are reported in online Appendix Table C4.

[21] These policy changes include the adoption, expansion, and expiration of the EUC program, as well as decisions by individual states to adopt optional EB triggers or make changes to their regular benefit programs.

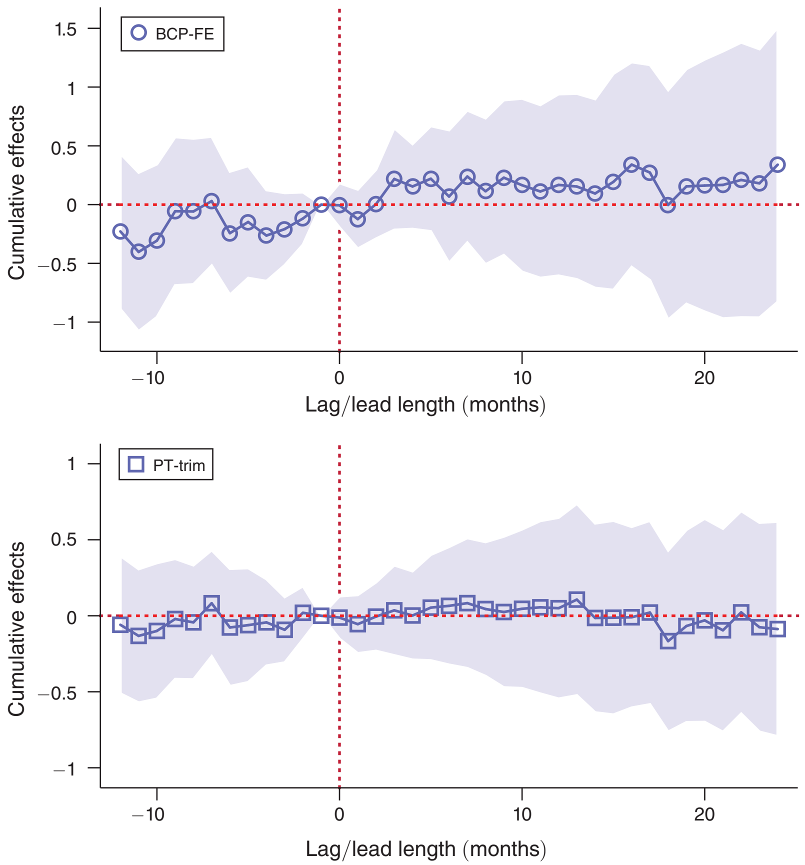

Figure 6. Cumulative Response of EPOP from Distributed Lags Specification: Full Sample Regressions

Notes: This figure reports the monthly cumulative response of EPOP from a 73-week increase in maximum UI benefit duration, centered around event date −1 whose cumulative response is defined as 0. The model is estimated on the full sample (2007:11–2014:12), using all border county pairs (BCPs) (hollow circles) and the subset of BCPs in the PT-trimmed sample (hollow squares), where all independent variables are divided by 73. The dependent variable is the first-differenced seasonally adjusted ratio of total employment to population age 15+, scaled in percentage points. The regression includes 24 lags and 11 leads in first-differenced benefit duration and is estimated using EPOP data from 2007:11–2014:12 (and thus duration data from 2005:11–2015:11). Lags are to the right of zero; leads are to the left of zero. The zeroth cumulative response is equal to the estimated coefficient on contemporaneous benefit duration. The *j*th cumulative lag is equal to the estimated coefficient on contemporaneous duration plus the sum of the estimated coefficient on the first through *j*th lag term. The *j*th cumulative lead is equal to −1 times the sum of the estimated coefficients on the first through the $(j-1)$th lead terms. The shaded region corresponds to the 95 percent confidence interval, robust to two-way clustering at the state and state-pair level.

mitigate the endogeneity problem, as discussed above in Section III. Nonetheless, to the extent that endogeneity bias remains, we can increase the probability of eliminating it by restricting our BCP analysis to exploit only the variation that is induced by national-level policy changes. Counties within a border pair are less likely to have systematically different employment trends when UI duration changes due to national policy than when one county's state is triggering on or off of EB or an EUC tier. We therefore develop an event study approach that isolates the effects of cross-border changes in benefit duration that are triggered by persistent changes in national policy and not by state-level economic shocks.

In addition, a recent econometric literature (e.g., Borusyak and Jaravel 2017, and Sun and Abraham, forthcoming) has shown that dynamic effects of treatment can confound lags and leads of treatment effects with cohort effects when the timing of treatment varies over the cross section and effects are heterogeneous. This is a potential concern with our full panel dynamic estimates since the timing of changes in UI generosity vary across counties even within border county pairs. The event study estimator we develop in this section of the paper uses cross-sectional differences in two separate national policy changes. Since the policy changes are simultaneous across all counties, our dynamic estimates are not confounded by cohort-specific heterogeneity in the treatment effects. As a result, our leads in the event study specification provide a cleaner test of preexisting trends.

The first policy change that we use is the passage of the Unemployment Compensation Extension Act (UCEA) in November of 2008, which granted states 20 weeks of federally funded benefits, or 33 if the total unemployment rate at the time exceeded 6 percent. This led to an increase in UI benefit durations that varied across states, introducing the first across-state variation in EUC availability in our sample.[22]

The second national policy change we use is the expiration of the EUC program in December 2013, which led to a reduction in UI duration that also varied across states. For example, benefits were reduced by 47 weeks in Illinois, Nevada, and Rhode Island but only by 14 weeks in Virginia, Iowa, New Hampshire, Minnesota, and 10 other states. Figure 2 shows a map of the reduction of UI duration at the end of 2013. As discussed above, North Carolina lost all EUC benefits, and the maximum benefit fell to 20 weeks a full 6 months before the national EUC expiration. As a result, we remove North Carolina from our baseline 2014 event study sample.[23]

Of course, the change in national policy creates variation precisely because there were differences in the *level* of unemployment across states. For the 2008 policy change, states that had a TUR exceeding 6 percent saw a bigger increase in benefit duration than did states with a lower TUR. Similarly, for the 2014 expiration, states with higher unemployment rates experienced larger reductions in benefits. While high and low unemployment states very well may have been on different trajectories around these two events, the BCP strategy is arguably better able to account for such trends compared to times when the policy change is directly induced by changes in state unemployment rates.

For our event study specifications, we use a two-year window—one year on each side of the national policy change. We regress EPOP on weeks of benefits, controlling for pair-period fixed effects and county fixed effects. We then instrument benefit duration with a variable that reflects only the change in duration caused by the EUC policy change. The instrument does not exploit variation caused by EB

[22] Prior to UCEA, variation in federally provided benefits existed in two states: North Carolina and Rhode Island were eligible for 13 and 20 weeks of EB, respectively, at the time of the policy change. No other state was eligible for EB at that time.

[23] To be clear, in the pooled estimates we discuss below, we include North Carolina in the 2007–2009 portion of the sample but exclude it from the 2013–2014 portion.

triggerings, EUC triggerings, and state-level policy changes. Our two-stage least squares estimation strategy is thus given by the set of equations

$$E_{cpt} = \beta D_{ct} + \lambda_c + \nu_{pt} + \eta_{cpt}, \tag{4}$$

$$D_{ct} = \beta_z z_{ct} + \rho_c + \gamma_{pt} + \epsilon_{cpt}, \tag{5}$$

where the instrument z_{ct} reflects the instantaneous change in the maximum UI duration available in the county due to the national EUC policy change. The instrument z_{ct} is defined as follows:

$$z_{ct} = \begin{cases} D_c^{08} & \text{November 2007–October 2008} \\ D_c^{08} + \delta_c^{08} & \text{November 2008–October 2009} \\ D_c^{13} & \text{January 2013–December 2013} \\ D_c^{13} - \delta_c^{13} & \text{January 2014–December 2014.} \end{cases}$$

For the 12 months prior to the 2014 policy change, we set the value of z_{ct} to equal the number of weeks of UI available in the last week of December 2013 (immediately prior to the EUC expiration), D_c^{13}. For the remaining 12 months in the sample, we subtract from D_c^{13} the number of weeks of benefits lost as a result of the EUC expiration (δ_c^{13}) and set z_{ct} equal to this value.[24] For the two-year window around the 2008 policy change, the instrument is defined analogously, using the maximum UI duration available just before (D_c^{08}) and just after the introduction of the new EUC program. Therefore, the jump in z_{ct} that occurs in November 2008 (δ_c^{08}) exactly equals the differential number of weeks made available by the onset of the UCEA. We also pool both events together and estimate this model using the 24 months of data around the 2008 onset along with the 24 months of data around the 2014 expiration.[25]

The instrumental variables (IV) estimates from equations (4) and (5) are presented in the bottom three panels of Table 2. In the second panel of Table 2, we report our pooled results using both the 2008 introduction (i.e., a positive treatment) and 2014 expiration (i.e., a negative treatment) of the EUC. For the PT-trimmed specification, the first-stage F-statistic for the excluded instrument is 262.3, indicating that the instantaneous changes due to the national policy changes were responsible for a sizable fraction of the variation in benefit duration over the event window; the first-stage coefficient is 0.842.[26] Our PT-trimmed second-stage estimate (0.253) is slightly larger than its PT-trimmed full sample counterpart, with a standard error of 0.650. While less precise than the full sample estimate, these

[24] Therefore, the change in the instrument z_{ct} between December 2013 and January 2014 takes into account the decline in duration explicitly resulting from the EUC expiration but not any contemporaneous changes in state-level regular benefits. In our robustness section, we show results from a specification where the instrument also takes into account the five state-level policy changes that occurred at the same time as the national policy change.

[25] For this pooled specification, we allow the county fixed effects to vary across the two subsamples (that is, the county fixed effects are replaced with county-by-subsample fixed effects).

[26] If the only changes in duration in the year before and the year after policy change were due to the policy change itself, the first-stage coefficient would be 1.

estimates using only national-level policy changes in the PT-trimmed sample can rule out employment reductions of -1.021 percentage points from the 73-week expansion of maximum benefit duration during the Great Recession. The point estimate from the untrimmed sample is similar (0.143) though less precise, with a standard error of 0.974. To the extent that the full sample specification contains some residual endogeneity that is purged in the event study specification, one would expect the former to produce more negative estimates than the latter. This is the case in the trimmed specification but not in the untrimmed specification. In any case, the differences between the full sample and pooled event study estimates are modest and insignificant.

The pooled estimates combine both the positive treatment in 2008 and the negative treatment in 2014. We also disaggregate the effects by time period. The 2008 results using the 2007:11 to 2009:10 period are reported in the third panel of Table 2, and the 2013:1–2014:12 period results are reported in the fourth panel.

The first-stage F-statistic on the excluded instrument for 2008 is substantially below the F-statistic for the pooled sample but is still above 40 for both the full sample and the PT-trimmed sample. The corresponding F-statistic for the 2014 expiration sample, by contrast, is 392.6 for the full sample and 423.8 for the PT-trimmed sample. The strength of the first stage in 2014 reflects the large size of the drop in duration upon expiration even within county pairs in 2014, the relative stability of duration in 2013, and the near-complete absence of changes in 2014. The 2014 expiration thus explains most of the variation in duration in the 2013–2014 time period. In contrast, though there was little within-pair variation in the year prior to the November 2008 expansion, duration changed substantially in the year following the expansion. In addition to a lower F-statistic, the first-stage coefficients are lower (near 0.7 in 2008 sample and over 0.9 in 2014 sample), reflecting the lower persistence of the initial duration change in 2008. Nonetheless, in both time periods, the instrument is strong.

In addition to greater policy persistence, the 2014 event also had lower EPOP variation. The 2008 onset was a time of great economic volatility. This is reflected in event study standard errors across the two samples. Standard errors in the full 2008 sample are 2.541; a 95 percent confidence interval is thus almost 5 percentage points of EPOP (two-thirds more than the decline in EPOP during the Great Recession in our sample). Even in the 2008 PT-trimmed sample, standard errors are 1.253. In contrast, the standard errors in the 2014 sample are between 0.5 and 0.6 for both the full and PT-trimmed samples. While substantially smaller than the 2008 standard errors, they are also larger than the standard errors from the full sample specification; this isn't surprising given the much smaller sample size and restricted variation.

The 2014 estimates are both very small and negative. The PT-trimmed estimate is -0.214, and results more negative than -1.239 can be ruled out with a 95 percent level of confidence. The PT-trimmed estimates in the 2008 sample are our largest in the paper. The point estimates imply that an increase of 73 weeks of UI duration increased EPOP by 1.344 percentage points. Thus, the estimates cannot rule out with a 95 percent level of confidence that a 73-week expansion reduced EPOP by less than -1.112 percentage points.

Dynamic Evidence.—We additionally show reduced form and first-stage estimates underlying the event study regressions by month relative to the event. As with the full sample regressions, the dynamic specification is estimated in first differences:[27]

$$(6) \qquad \Delta E_{cpt} = \sum_{\tau=-12}^{11} \beta_\tau \delta_{ct} \mathbf{1}\{eventdate_t = \tau\} + \nu_{pt} + \eta_{cpt},$$

$$(7) \qquad \Delta D_{ct} = \sum_{\tau=-12}^{11} \beta_{z\tau} \delta_{ct} \mathbf{1}\{eventdate_t = \tau\} + \gamma_{pt} + \epsilon_{cpt}.$$

We let $\delta_{ct} = \delta_c^{08}$ for the 2007–2008 sample and $-\delta_c^{13}$ for the 2013–2014 sample, each divided by 10 for ease of interpretation. The sum of coefficients $\rho_\tau = \sum_{k=-11}^{\tau} \beta_k$ and $\rho_{z\tau} = \sum_{k=-11}^{\tau} \beta_{zk}$ represent the cumulative response by event time. These represent the average within-pair differences in employment and the prevailing maximum benefit duration—over a 24-month window around the national policy change—for a pair in which the difference in the instantaneous increase in maximum benefit duration (due to the policy change) was 10 weeks. We omit the variable corresponding to $eventdate_t = -1$ (which corresponds to October 2008 and December 2013), meaning that the plotted coefficients are centered relative to date -1 leading values.[28]

Figure 7 shows these first-stage and reduced-form estimates period by period around the event date as compared to the values from the month just prior to treatment (i.e., -1). The EPOP difference between the two sides of the border is plotted on the left-hand *y*-axis, with the difference in maximum benefit duration plotted using the right-hand *y*-axis. Panel A displays the pooled estimates, while panels B and C show results for the 2008 and 2014 events separately. The corresponding results for the PT-trimmed sample are displayed in online Appendix Figure C4. The dynamic evidence mirrors the numerical results in Table 2. In the pooled sample, treatment in the 2014 period is defined positively in both the reduced form and first stage. This does not alter the sign of the event study estimates but does make the first stage positive going forward in time to line up with the 2008 period. The pooled estimates show (by construction) a clear increase at date 0 of approximately 10 weeks in the maximum benefit duration relative to the neighboring county.[29] Much of this increase in benefits persists over the following 12 months. There is little indication of a differential trend in EPOP prior to the national-level policy changes, which provides additional validation for the event study coupled with the border county design. Importantly, employment relative to population remains fairly stable over the 12 months following treatment; we see little indication of job loss following the

[27] We note that estimating this model in levels (i.e., using E_{cpt} and D_{ct} and mean differencing) versus first differences is immaterial in this case, where we are estimating monthly coefficients, β_τ, over a fixed 24-month sample. Estimating the model in levels yields numerically identical estimates.

[28] For ease of interpretation, we omit January 2014 instead of December 2013 in the first stage when constructing the graph that analyzes only the 2014 expiration event. This allows the graph to show a drop in relative benefits roughly from 10 to 0 rather than 0 to -10. As we do not report standard errors for this specification, this amounts to a simple vertical shift of the graph.

[29] The increase is not exactly ten weeks, because the policy changes in question did not occur precisely at the end of a calendar month.

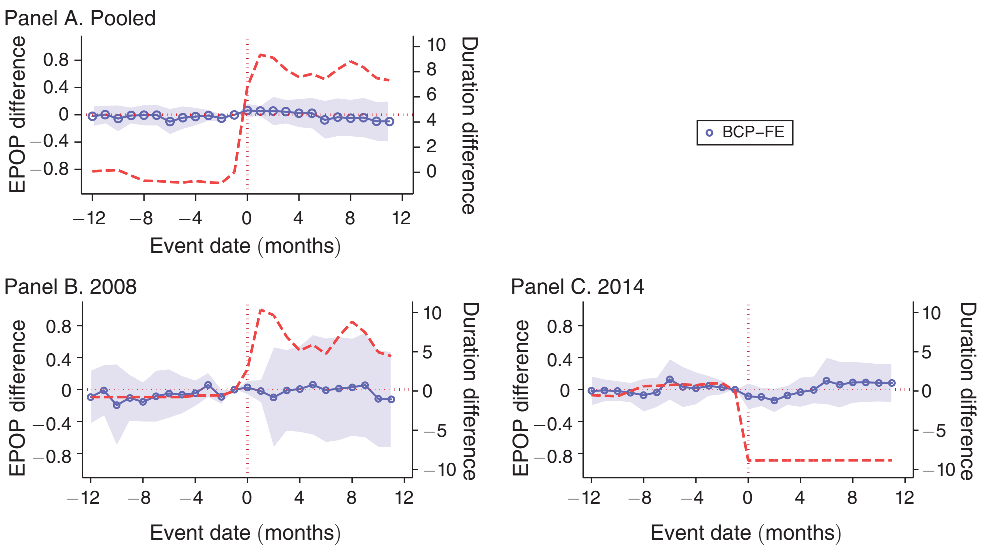

Figure 7. Evolution of EPOP Difference and UI Benefit Duration Difference across State Borders: 2008 Expansion and 2014 Expiration of EUC

Notes: This figure reports the monthly cumulative response of EPOP (left axis, hollow circles) around the 2008 expansion and 2014 expiration of cross-state differentials in UI benefits. Panel A uses pooled 2008 and 2014 samples centered around event date −1 whose cumulative response is defined as zero. Panels B and C separately examine the 2008 and 2014 events. The dependent variable is the first-differenced seasonally adjusted ratio of total employment to population age 15+, scaled in percentage points. The regression includes 11 lags and 12 leads in first-differenced benefit duration: for the 2008 sample, the duration variable is equal to the increase in weeks of UI duration immediately upon the implementation of UCEA, divided by 10; for the 2014 sample, the duration variable is defined as −1 times the weeks of UI duration lost as a result of EUC expiration, divided by 10. The dashed line (right axis) reports the monthly cumulative response of benefit duration around the event; the regression is identical to the EPOP specification except that the dependent variable is the first-differenced benefit duration in weeks. Event date zero is marked with a dotted vertical line; this corresponds to November 2008 for the 2008 sample and January 2014 for the 2014 sample.

national-level policy changes. Furthermore, the results are visually similar in both the baseline BCP-FE and the refined PT-trimmed sample.

In the 2008 sample, there is also little indication of systematic employment changes—either in the year prior to the 2008 UCEA implementation or during the subsequent year. As with the results in Table 2, the estimates for 2008 are less precise. In addition, the figure shows that the duration differences are somewhat less persistent. Overall, while noisy, the estimates from the 2008 event (especially from the more precise trimmed sample) are broadly consistent with those from the pooled estimates and do not indicate substantial losses in employment from this policy change. Finally, the 2014 figure does not show much of an effect on EPOP from the program expiration. The duration differences between county pairs were much more persistent (looking backward in time) compared to 2008, mostly exceeding 80 percent of their immediate preexpiration duration during the entirety of 2013. This explains why the first-stage coefficient is much closer to unity: 0.915 for the baseline BCP sample and 0.903 for the PT-trimmed sample.

Overall, both the full sample and event study estimates suggest that there was no sizable positive or negative employment effect of the 73-week increase in UI

Table 3—Robustness of the Effects of UI Benefit Duration on EPOP: Choice of Sample Period

	BCP-FE (1)	PT-Trimmed (2)
1. 2007:11–2014:12	0.430	0.180
	(0.471)	(0.268)
Observations	199,692	149,640
2. 2006:11–2014:12	0.142	0.107
	(0.455)	(0.272)
Observations	227,556	170,520
3. 2005:11–2014:12	−0.088	0.030
	(0.445)	(0.277)
Observations	255,420	191,400
4. 2004:11–2014:12	−0.330	−0.062
	(0.457)	(0.286)
Observations	283,284	212,280
County pairs	1,161	870

Notes: Each cell reports the coefficient on D_{ct} from a regression of the form $E_{cpt} = \beta D_{ct} + \lambda_c + \nu_{pt} + \eta_{cpt}$. E_{cpt} is the seasonally adjusted ratio of total employment to population age 15+, scaled in percentage points, and D_{ct} is the potential weeks of UI benefits divided by 73. The second column restricts the sample to the PT sample. PT-trimming removes the quartile of county pairs with the highest mean squared error in EPOP between November 2004 and October 2008 (after partialling out a fixed level difference). The regression in each row is estimated over the sample period indicated. The estimates in row 1 correspond to the estimates in panel A of Table 2. Standard errors are two-way clustered at the state and state-pair level.

maximum duration during the Great Recession. This is true when we use all policy variation in our full sample specifications or when we instrument the policy variation using national-level changes. Our dynamic evidence suggests no employment changes for the first year and a half following the policy innovations. When we consider the refined BCP strategy that excludes some of the more poorly matched pairs, we find no evidence of employment changes up to 24 months following treatment.

C. *Robustness of Estimates*

Choice of Sample Period.—Table 3 shows results from the full sample specification for alternative sample periods beginning in 2007:11, 2006:11, 2005:11, and 2004:11.[30] The first column shows results for the baseline BCP sample, and the second column shows results for the PT-trimmed sample. Overall, the baseline BCP estimates range between 0.430 and −0.330, while the PT-trimmed estimates range between 0.180 and −0.062. Importantly, while the estimates differ in size, we stress that none of the eight estimates shown in Table 3 are statistically significantly different from zero at conventional levels.

[30] By pushing the start date further back in time, we are only adding data from the pretreatment period; there is essentially no variation in UI benefits between 2004 and 2007.

At the same time, the baseline BCP estimates vary somewhat by sample, and these estimates decrease monotonically in the length of the window: the earlier the sample start date, the more negative the estimate. The gap between the estimate for the sample starting in November 2007 and the sample starting in November 2004 is nontrivial; it represents a differential impact of 0.76 percentage points of EPOP from a 73-week increase in UI duration. This is almost twice the baseline estimate starting in 2007:11. Note that the pattern in the estimated effect is consistent with the presence of a downward trend in EPOP in treatment counties relative to control. As we discussed above, and as shown in Figure 4, we see a slight relative decline in EPOP throughout the 2004–2014 period on the side of the border that is more heavily treated during the treatment period. Adding observations from a time period when EPOP was relatively higher on the high-treatment side and when treatment was low mechanically makes the estimated treatment effect more negative. These differential trends could be consistent with some degree of residual endogeneity or with serially correlated noise. Regardless of the source of these differential trends, a 2007:11–2014:12 sample frame—with 12 months before treatment begins and after treatment ends—ensures that any differential trends between counties is approximately orthogonal to D, our independent variable of interest. By contrast, with a larger amount of time before treatment than after treatment, these trends are no longer orthogonal to D, potentially leading to bias.

The variation in estimates is much smaller for the PT-trimmed estimates: the 2007–2014 estimate is 0.180, and the 2004–2014 estimate is -0.062. This mechanically reflects the fact that the magnitude of differential trends is much smaller in PT-trimmed sample, though we cannot say conclusively whether this primarily reflects a reduction in noise or in residual endogeneity.

Trimming on Pretreatment Mean-Squared Error.—The refined BCP strategy trims the pairs with the worst matches—25 percent of the sample with the biggest out-of-sample mean squared prediction error in pretreatment EPOP. In online Appendix Table C5, we show how our four main estimates (full sample, 2008 event study, 2014 event study, and pooled event study) vary as our threshold for trimming varies. We show estimates for different trimming thresholds across seven rows. The rows are, respectively: no trimming, 10 percent trimming, 20 percent trimming, 25 percent trimming, 30 percent trimming, 40 percent trimming, and trimming at the median of the difference in pretreatment MSE. The 25 percent trim is our main PT-trimmed specification from Table 2. The coefficient estimates are fairly robust to changes in the trimming threshold. The standard error is minimized for the full sample at a 25 percent trim. It is minimized at a 50 percent trim for the pooled event study sample, 40 percent for the 2008 sample, and 10 percent for the 2014 sample. Thus, our choice of a 25 percent benchmark trim across all specification is a reasonable one.[31]

Additionally, for all specifications, the primary impact of trimming on the coefficient estimates seems to be a reduction in the standard errors by reducing residual

[31] Note that after dropping the 10 percent of county pairs with the worst matches, standard errors remain relatively stable with further trimming.

variation. It does not seem to systematically change the magnitude of the estimate in a positive or negative direction. The reduction in the standard errors is often up to 50 percent from the baseline sample. The one exception is the 2014 estimate, where the maximum reduction across trimming thresholds is approximately 20 percent.

Additional Robustness Checks.—In Table 4, we consider a number of other robustness checks for our estimates on the full 2007–2014 sample and for our pooled event study. We do this for both the baseline BCP sample and the PT-trimmed sample. The first row in the table reproduces the estimates from Table 2. Each of the remaining rows varies the specification, data, or sample as follows. We show estimates of impacts on private employment only. As an additional strategy to mitigate residual mechanical endogeneity, we drop pairs containing counties that show a high correlation between county EPOP and the EPOP of its state over the 2004:11–2008:10 period ("correlation trimming").[32] Comparison within these county pairs should be less prone to contamination from state-specific employment shocks that endogenously determine state-level benefit duration. We include an (in-sample) county specific linear trend (ISLT) control. We trim based on pretreatment MSE estimated over the 2004:11–2007:10 period (instead of 2004:11–2008:10). Because the temporary lapses in EUC extensions in the absence of Congressional reauthorization (correctly) might not have been seen as changes, because they were expected to be reversed in a very short period of time, we recode treatment during these lapses at the level of the duration during the last week before the lapses; we do not recode for the event study estimates because none of the lapses occur during the relevant sample periods. We also estimate using quarterly as opposed to monthly data: once using the same QCEW employment data but aggregated to the quarterly level and once using quarterly employment statistics from a different dataset, the Quarterly Workforce Indicators (QWI, US Census Bureau 2016b). We show results using data that have not been seasonally adjusted. We also estimate a specification where we allow for imbalance in our panel by including counties with missing values in the sample. We additionally use a log-log specification instead of the level-on-level specification used throughout the paper. We do this using both log employment and log EPOP as outcomes but also report the EPOP-equivalent estimates in square brackets for comparability.[33]

Next we show a pooled event study specification where we instrument using the total change in benefits rather than the change in benefits due solely to the expiration of EUC. In this case, the instrument includes the additional decreases below 26 weeks made by state governments in Florida, Georgia, Kansas, and South Carolina, as well as an increase from 26 to 30 weeks in Massachusetts. We also show three different specifications where we alter our baseline treatment of North Carolina, which lost access to EUC benefits earlier than other

[32] This is motivated by comments on HKMM by Hall (2013).

[33] For instance, the estimate of 0.005 in column 2 for log EPOP would imply that the expansion of UI from 26 to 99 weeks increased EPOP by $\left((99/26)^{0.005} - 1\right) \times 42 = 0.282$ percentage points (since the unweighted mean EPOP in this sample is approximately 42), similar to the coefficients that we see in the level-on-level specification (0.180). The level equivalents for the log-log specification are displayed in brackets below the coefficient estimates. The level-on-level equivalents of the log employment estimates are quite close to the original estimates.

Table 4—Additional Robustness Checks on the Effects of UI Benefit Duration on EPOP

	Full sample OLS		Pooled sample IV	
	BCP-FE (1)	PT-Trimmed (2)	BCP-FE (3)	PT-Trimmed (4)
1. Baseline	0.430	0.180	0.143	0.253
	(0.471)	(0.268)	(0.974)	(0.650)
2. Private EPOP	0.268	0.010	0.205	0.389
	(0.495)	(0.264)	(1.023)	(0.647)
3. Correlation-trimmed	−0.142	−0.007	−0.120	0.106
	(0.354)	(0.269)	(1.118)	(0.644)
4. ISLT	0.380	0.117	0.930	0.756
	(0.369)	(0.226)	(0.680)	(0.336)
5. Eliminate lapse	0.543	0.210		
	(0.521)	(0.279)		
6. Quarterly data	0.453	0.196	0.205	0.303
	(0.512)	(0.292)	(0.928)	(0.634)
7. QWI EPOP (quarterly)	0.692	0.495	0.402	0.556
	(0.481)	(0.328)	(0.646)	(0.614)
8. Not seasonally adjusted	0.301	0.146		
	(0.486)	(0.274)		
9. Unbalanced panel	0.329	0.180	0.148	0.253
	(0.474)	(0.268)	(0.957)	(0.650)
10. ln(*EPOP*)	0.006	0.005	0.007	−0.001
	(0.008)	(0.007)	(0.017)	(0.012)
	[0.335]	[0.312]	[0.424]	[−0.068]
11. ln(*employment*)	0.008	0.007	0.014	0.004
	(0.009)	(0.009)	(0.016)	(0.012)
	[0.429]	[0.387]	[0.802]	[0.253]
12. Exploit Δ reg. benefits			0.185	0.265
			(0.960)	(0.628)
13. Drop NC	0.416	0.137	0.159	0.264
	(0.556)	(0.318)	(1.005)	(0.674)
14. Keep NC			−0.147	−0.046
			(1.044)	(0.745)
15. NC: Alt. instrument			0.071	0.120
			(0.629)	(0.436)
16. Distance trimming	0.323	0.241	0.313	0.488
	(0.406)	(0.277)	(1.131)	(0.710)
17. Hinterland pairs	0.939	0.841	0.041	−0.313
	(0.539)	(0.411)	(1.064)	(0.691)

Notes: Each cell reports regressions analogous to those reported in Table 2 for the full sample with OLS or the pooled event samples (IV). The estimates in the first row correspond to the estimates in panels A and B of Table 2. The estimates in the second row replace (total) EPOP with the ratio of private employment to population age 15+. In the third row, we trim the set of border county pairs based on the level of correlation between county EPOP and state EPOP over the period 2004:11–2008:10 (see text for details). The fourth row controls for county-specific linear trends. The fifth row recodes the periods in 2010 when EUC lapsed by assigning EUC values during these lapses as equal to their prior value. The sixth row uses quarterly data instead of monthly (and estimates over the 2007:IV–2014:IV period). The seventh row uses EPOP derived from the QWI (at the quarterly level) instead of the QCEW. The eighth row uses seasonally unadjusted data. The ninth row includes counties without full EPOP data for each month, which we drop by default. The tenth and eleventh rows use ln(*EPOP*) and ln(*employment*), respectively, as dependent variables. The bracketed estimates in these two rows are the level-on-level equivalent, equal to $((99/26)^{\hat{\beta}} - 1)\bar{E}$, where $\bar{E}$ is the mean EPOP level in the given sample. The twelfth row uses a modified version of the instrument z_{ct} that exploits all changes in benefits, including changes in regular benefits, which occur at the end of December 2013. Rows 13–15 report estimates using alternative strategies for dealing with North Carolina (NC); by default, border county pairs (BCPs) with one neighbor in NC are kept in the full sample OLS and the 2008 subsample and are dropped in the 2014 subsample. The thirteenth row completely drops all NC BCPs. The fourteenth row keeps all North Carolina BCPs. The fifteenth row keeps NC BCPs but redefines the instrument for NC counties (see text for details). The sixteenth row drops county pairs whose population centroids are greater than 100 kilometers apart. The seventeenth row uses the "hinterland" pairs rather than the border pairs (see text for details). Cells that are not applicable in the given sample or that provide estimates that are mechanically equal to the baseline estimates are left blank. Standard errors are two-way clustered at the state and state-pair level.

states.[34] In the sixteenth row, to demonstrate that our controls are well-matched to our treatments, we repeat the main estimates when restricting the sample to a plausibly better-matched group of pairs whose population centroids are less than 100 kilometers apart.[35]

For the full sample specifications, the lowest estimates are the correlation trimmed estimates at −0.142 for the baseline sample (column 1) and −0.007 for the PT-trimmed sample (column 2). The highest estimates (not including row 17, which we will discuss below) are the quarterly QWI estimates at 0.692 for the full sample and 0.495 for the trimmed sample. Thus, the variation in the estimates is relatively small. For the event study specifications in columns 3 and 4, the estimates range between −0.147 and 0.930 for the baseline BCP sample and between −0.046 and 0.756 for the PT-trimmed sample. The standard errors are generally larger for the event study estimates, as expected. Overall, across these full sample and event study estimates, 48 out of 54 are positive, and none are below −0.15. Only 1 of the 54 estimates (+0.756) is statistically significant with a 95 percent or greater level of confidence. If each estimate were an independent random draw under the null, we would expect to see at least one significant coefficient by chance 93.7 percent of the time. In sum, our estimates show only modest variation, very few are negative, and none of the negative estimates are statistically significant.

In online Appendix Table C6, we show the robustness checks for the 2008 and 2014 event study analyses separately. The results are largely similar to our pooled event study results, though the standard errors are significantly larger for the 2008 event study and often 30–50 percent smaller for the 2014 event study. The 2008 event study estimates are imprecise because the initial 2008 triggering explains less of the variation in treatment in the surrounding two-year sample period. In addition, they are imprecise because of the large variation in EPOP during the onset of the Great Recession.

Robustness to Spatial Regression Discontinuity Controls.—In their critique of Hagedorn et al. (2019), Dieterle, Bartalotti, and Brummet (2020) argue that border county pairs are too coarse as geographic controls. They instead recommend using a regression discontinuity estimate, effectively relying upon counties whose population are located very close to the state border. Specifically, they show that adding spatial distance controls to the border county design reduces the magnitude of the HKMM estimates to near zero and renders them statistically insignificant. Dieterle, Bartalotti, and Brummet (2020) implement their design using county-level data by computing the population-weighted average distance to the border for every county

[34] Recall that North Carolina lost access to EUC at the end of June 2013. This was a full six months before the other states lost access to EUC benefits, which means that North Carolina gets treated halfway through the control period in the 2014 event study analysis. In our main specifications analyzing the 2014 EUC expiration, therefore, we drop all county pairs containing a county from North Carolina; we also drop North Carolina from the 2014 part of the sample in the pooled event study regression. As robustness checks, we drop North Carolina from the entire baseline BCP-FE full sample estimation as well as from the entire pooled event study specification (row 13). We next include North Carolina in the 2014 portion of the pooled event study specification (row 14). Finally, we retain the inclusion of North Carolina in the 2014 portion of the pooled event study sample but redefine the instrument, in North Carolina's case, to reflect the drop in EUC benefits for North Carolina in July 2013 (row 15).

[35] We discuss row 17 in the following section.

along each state border segment. The addition of spatial controls focuses implicitly on counties located very close to the border.

We show in a number of ways that the concerns expressed in Dieterle, Bartalotti, and Brummet (2020) about poor matches are not warranted in our case. Above, we showed both covariate balance (in Table 1) and absence of differential trends within county pairs prior to treatment (in Figures 4 and 6). In this section, we demonstrate that our findings are robust to the inclusion of the spatial distance controls prescribed by Dieterle, Bartalotti, and Brummet (2020). Online Appendix Table C7 displays the results from application of this RD estimator using EPOP as the dependent variable. In their estimation, Dieterle, Bartalotti, and Brummet (2020) include pair-period fixed effects but omit county fixed effects.[36] A priori, inclusion of county fixed effects is important if the results are to have a difference-in-difference- (or difference-in-discontinuity-)type interpretation. As a practical matter, we show that the omission of county fixed effects yields uninformative estimates. With EPOP as the dependent variable and no county fixed effects, the Dieterle, Bartalotti, and Brummet (2020) estimator implies that raising UI benefits from 26 to 99 weeks led to a contraction in EPOP of 27.247 percentage points, as reported in the top row of the second column of online Appendix Table C7. The standard errors are even larger at 35.819, rendering these estimates sufficiently imprecise as to be not useful. Importantly, adding county fixed effects to the Dieterle, Bartalotti, and Brummet (2020) specification dramatically lowers the magnitude of the estimates as well as the standard error, as reported in columns 3 and 4. With county fixed effects and no spatial controls (column 3), the estimate falls in magnitude from -27.247 to 0.303, while the standard error drops from 35.819 to 0.300. Conditional on county fixed effects, our BCP-FE estimates are highly robust. The addition of spatial RD controls (column 4) does increase the standard errors, but it has no substantive impacts on our mean BCP-FE estimates: the estimate increases to 0.612, and the standard error roughly doubles to 1.155. The inclusion of county fixed effects in addition to pair-specific period fixed effects is critical to the BCP-FE research design, as shown by this exercise.[37]

Though it is reassuring that our BCP-FE estimates are robust to the addition of spatial controls, the Dieterle, Bartalotti, and Brummet (2020) estimator does yield wider standard errors than our BCP-FE estimator does. This is to be expected since, effectively, the Dieterle, Bartalotti, and Brummet (2020) estimator is estimating off of a substantially smaller set of counties that are particularly geographically close. There are nonborder ("hinterland") counties that are close to the border, and we can add Hinterland County Pairs (HCP) to our estimation to gain precision when

[36] Dieterle, Bartalotti, and Brummet (2020) use state-pair-by-period fixed effects in their regression. In our baseline specification, replacing (county-)pair-period fixed effects with state-pair-by-period fixed effects would mechanically have no effect. The specification used by Dieterle, Bartalotti, and Brummet (2020) is slightly different, making the use of state-pair-by-period fixed effects appropriate.

[37] Note that there are other more minor differences between our specification and that of Dieterle, Bartalotti, and Brummet (2020) that account for small differences in estimates beyond the very large differences due to omission of county fixed effects in the Dieterle, Bartalotti, and Brummet (2020) model. In particular, each county lies only in one county pair in Dieterle, Bartalotti, and Brummet (2020), in contrast to our setting where a county is in one pair for each county in another state that it borders. The (small) effect of these specification differences are illustrated in online Appendix Table C8.

using the regression discontinuity design. Intuitively, they enable greater precision in estimation of the gradient leading up to the border. We form the set of hinterland county pairs by considering all counties that are adjacent to border counties but are not border counties themselves; two hinterland counties are paired if they each border a member of the same border county pair.[38] The addition of the HCP counties to the RD specification yields very similar point estimates (0.553) but almost halves the standard errors. We show this in the second row of online Appendix Table C7.

The HCP sample is also helpful in addressing another possible limitation of the BCP design: policy-induced spillovers across the border. Because the counties in the HCP sample are not adjacent to each other, such spillovers should be much lower than they are for the BCP sample. There is, of course, a trade-off here: while the geographic spillovers should be greatly reduced, the hinterland counties may not be as good controls for each other, which could lead to estimates that are less precise or more affected by reverse causality. Row 17 of Table 4 displays the HCP estimates. The event study estimates (columns 3 and 4) are largely unchanged. The full sample estimates (columns 1 and 2) are somewhat more positive than the BCP-FE estimates: the baseline HCP estimate is 0.939, while the PT-trimmed HCP estimate is 0.841, the latter being statistically significant at the 95 percent confidence level. We explicitly test for and cannot reject at conventional levels the hypothesis that the two sets of estimates (BCP and HCP) are the same.

Additionally, in online Appendix Table C9, we interact our treatment variable (duration) with the distance between the two members of each pair (centroid to centroid) in bins. In column 1, we restrict the sample to the set of border county pairs; in column 2, we restrict to the set of hinterland county pairs; and in the third column, we pool both sets of pairs together. The coefficient in the first row represents the "main effect"—i.e., the effect for the omitted category, which is the set of pairs with the smallest intercentroid distances. The remainder of the coefficients represent the effect in each respective bin relative to those small-distance pairs. In column 2, there is some evidence that the effect is slightly more negative as the distances increase; however, this pattern is not present in either column 1 or 3. In all cases, a test for the significance of the set of interaction terms does not reject that they are jointly zero. Moreover, estimates from the BCP + HCP sample are narrow given the small sample sizes, ranging from -0.497 to $+0.654$. Overall, we find little evidence that either endogeneity, due to poor matches between treatment and control, or biases, due to cross-border spillovers, are important in driving our findings.

D. *External Validity: Size and Persistence of Policy Changes*

One potential concern with our border-county-pair design is whether the differences in UI benefit duration between counties across the state border were sizable and persistent, especially as compared to the national-level changes in benefit

[38] For example, consider Broome County (NY) and Susquehanna County (PA), which are adjacent to each other and thus contained in our BCP sample. Cortland County (NY), which is adjacent to Broome, is located in the "hinterland" of NY—that is, in the interior of the state and not along the border. Likewise, Lackawanna County (PA) is adjacent to Susquehanna and not on the PA border. Our HCP sample would therefore include the hinterland pair Cortland (NY)–Lackawanna (PA).

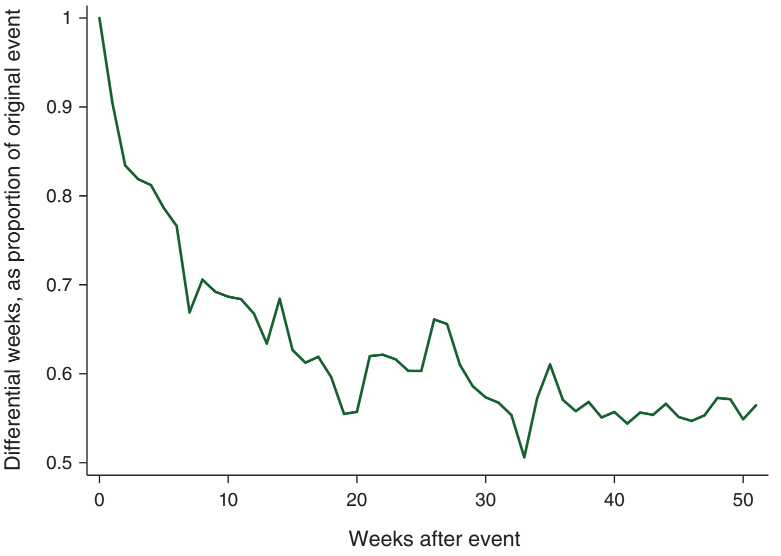

Figure 8. Persistence of Differential Change in UI Benefit Duration across Border County Pairs

Notes: This figure plots the persistence of all changes in relative duration in the full sample. In particular, the data is organized at the pair (p), event (s), event-week (τ) level, where an event is any change in the duration difference across a county pair. The dependent variable $y_{ps\tau}$ is the difference in duration across the county pair minus that same difference immediately prior to the event. This dependent variable is regressed on the size of the initial event interacted with 52 dummies for the 52 event-weeks τ immediately following the event. This figure plots those coefficients. See text for details.

duration that took place during the Great Recession. Online Appendix Figure C5 shows the distribution of differences in maximum benefit duration across county pairs and over time for the full sample. Here, each observation is a county pair in a given week between November 23, 2008, and December 22, 2013. As the figure shows, around 40 percent of pair-week observations in this sample have no difference in UI benefit duration, while nearly half of the observations have a benefit duration exceeding 10 weeks. To put this in perspective, a 10-week differential is almost 40 percent of the typical maximum benefit duration of 26 weeks that prevailed in all but two states prior to the Great Recession. Therefore, the gaps across state borders that we are evaluating are economically substantial. In online Appendix Figure C6, we show that similarly sized duration gaps existed between the two sides of the border just prior to the EUC expiration in 2014.

The gaps in UI benefit duration between neighboring counties across the border were substantial, but were they also persistent? Figure 8 shows the mean benefit duration gap (as a share of the initial gap) by weeks following a particular event.[39] On average, 10 weeks after the event, 70 percent of the original gap in maximum

[39] In this analysis, all changes in relative benefit differences are treated as "events" or "shocks." With the data organized at the pair-by-shock (ps) level, we regress the change in relative duration on a set of $shock_{ps} \times eventdate_{\tau}$ indicator variables, where $shock_{ps}$ is the size of the initial shock and $eventdate_{\tau}$ runs from 0 to 51 weeks after the initial shock. For instance, suppose at time t, county A increased duration from 53 to 63 weeks while county B held constant at 47 weeks; then $shock_{ps}$ would be equal to 10. The dependent variable in the regression (for $\tau = 0, 1, \ldots, 51$)

benefit duration between the 2 sides of the border remained in place. Even 52 weeks after the event, on average, more than 50 percent of the original gap in duration persisted across the border. Overall, the evidence suggests that the benefit durations we are using for identification are not transitory policy shocks. The duration series in Figure 7 shows similar information for the specific 2008 and 2014 events.

We additionally show that the high average persistence of the policy shocks is not driven by a small number of cases but rather that policy persistence was widespread across counties. In panel A of online Appendix Figure C7, we show the share of counties where the duration gap continuously remained at least as large as the initial gap by weeks following the the 2008 event. The figure shows that after approximately 20 weeks, the initial gap remained in place or increased in about 60 percent of the county pairs; by 40 weeks, about 15 percent of the pairs retained the full gap. Panel B shows evidence for the 2014 expiration, looking backwards in time. Even 50 weeks before the EUC expiration, over 40 percent of counties had gaps in duration at least as large as the gap at the time of expiration. Thus, the 2014 event study estimates are based on the expiration of highly persistent differentials across county pairs.

Overall, while the cross-sectional differences in size and persistence of the UI benefit duration are not as dramatic as the overall national-level changes that occurred during the Great Recession, they are nonetheless quite substantial—especially for the 2014 expiration event. Moreover, the persistence of the events in our samples are quite a bit greater than those used in some of the other papers in the literature. For their main specification, Chodorow-Reich, Coglianese, and Karabarbounis (2019) use treatment events whose half-life is roughly eight weeks (see their Figure 2). In contrast, as shown in our Figure 8, the half-life of the typical event used for our baseline full sample estimate exceeds 52 weeks.

V. HKMM and HMM Reconciliation

In this section, we provide a brief reconciliation of our estimation results with those of HKMM and HMM. We do this because we use similar methods but end up with results that are quite different. We first compare our full sample results to those in HKMM and then compare our event study results to those in HMM. Our replication of the HKMM estimates, joint with Dieterle, Bartalotti, and Brummet (2020), is discussed in online Appendix D. Additionally, we expand upon this section in further detail in online Appendix B.[40]

would be equal to $D_{A,t+\tau} - D_{B,t+\tau} - 6$, since the preshock difference was 6 weeks. Therefore, the regression coefficients trace out the share of the original shock that remains after τ weeks.

[40] Since our original joint replication exercise with Dieterle, Bartalotti, and Brummet (2020), HKMM released a new version of their paper (Hagedorn et al. 2019). Their estimates of the impact of log UI duration on log EPOP increased slightly from 0.049 to 0.053, while the coefficient in our replication is 0.051. Since the differences in their estimates are small and both are close to our replication of their estimates, we have maintained our decompositions relative to our replication in the joint online Appendix.

A. *HKMM Comparison*

The point estimate from our baseline (non-PT-trimmed) specification suggests that an expansion of UI benefit duration from 26 to 99 weeks raises the employment-to-population ratio by 0.430 percentage points. In comparison, our replication of the HKMM specification suggests that the same expansion in duration *reduces* the employment-to-population ratio by 2.661 percentage points.[41] The differences between these specifications can be decomposed into six distinct choices—three of which are consequential and three of which are not. The three relatively nonconsequential choices are: (i) we use levels of variables, whereas HKMM use logs; (ii) we control for county fixed effects and time fixed effects in our model, whereas HKMM use interactive fixed effects; and (iii) we eliminate four counties with gaps in reporting in the QCEW during our sample period, while HKMM use an unbalanced sample. The three consequential choices are: (iv) HKMM quasi-forward difference their dependent variable, whereas we do not; (v) we use employment from the QCEW to construct our main dependent variable, whereas HKMM use unemployment from LAUS; and (vi) we use a symmetric one-year window surrounding the time period of differential UI expansion (2007:11–2014:12), whereas HKMM use the time period 2005:1–2012:12.

These six differences mean that transitioning from one specification to the other involves a series of six distinct steps. We compute all of the permissible transition paths between the HKMM specification and ours (and we summarize the results in online Appendix Table B1). We characterize the impact of each choice by computing the average effect of making each individual change across all permissible combinations of the five other assumptions.[42] In other words, we separately estimate the average marginal impact of moving from HKMM's estimates to ours across all configurations of assumptions for each other choice.[43]

The total gap between our baseline non-PT-trimmed estimate and our replication of the HKMM estimates is 3.091 percentage points of the employment-to-population ratio. Out of this, averaged across specifications, 42 percent is due to the use of quasi-forward differencing, 30 percent is due to alignment of the sample period, and 22 percent is due to the usage of LAUS data as opposed to QCEW data (column 3 of online Appendix Table B1). The marginal effects for the first two of these three changes are statistically significant at the 99 percent confidence level, while the

[41] The HKMM dependent variable is (a function of) the unemployment rate, while ours is EPOP. We translate implied effects on the unemployment rate into implied EPOP effects by scaling each by their relative drops over the course of the Great Recession.

[42] Since quasi-forward differencing only makes sense with logs as opposed to levels, we do not consider the marginal impact of moving from logs to levels in specifications with quasi-forward differencing. Since there are 6! different sets of choices but half of them contain quasi-forward differencing and variable levels, we have in total $6!/2$, or 360, usable combinations. We compute standard errors for our estimates of the average marginal effects for each change using a state-level block bootstrap with 200 replications.

[43] These average marginal impacts are displayed in column 3 of online Appendix Table B1. The table also presents two other ways of characterizing the relative impacts of these specification choices as we move between HKMM's specification and ours. Column 1 shows the effect on HKMM's estimates of separately implementing each change, while column 2 shows the effect of taking each final step to arrive at our specification. No estimates are displayed for the cells corresponding to a quasi-forward differenced levels specification. The one-off changes from both the HKMM estimate and our estimate are larger than the averages along the transition path, reflecting that the various differences between the two estimates are substitutes.

third is significant at the 90 percent confidence level. The average marginal effect of the other three changes are substantially smaller, with t-statistics well below 1. The use of interactive as opposed to linear fixed effects account for only 6 percent of the difference, and the use of logs as opposed to levels as well as the alignment of the county sample each explains less than 0.3 percent of the difference. For these latter three changes, there are reasonable arguments for either decision, and we view them as robustness checks. It is therefore encouraging that our results are not very sensitive to these choices.

For each of the first three relatively consequential changes, however, we argue that our specification choices are preferred. HKMM use quasi-forward differencing in order to deal with policy anticipation. As we show in Figure 6, there is no detectable policy anticipation up to even one year in advance, suggesting that the problem that QFD is designed to solve is not present. Moreover, the use of QFD introduces several additional problems. First, with a dependent variable in QFD form, it is difficult to separate out policy anticipation in period t from a standard Keynesian effect in period $t + 1$. Second, as shown in online Appendix Figure B1, QFD mechanically converts a (small) negative lagged effect on unemployment into a (large) positive effect and can thus can introduce a large bias. The second consequential choice is the use of LAUS data to measure unemployment as opposed to QCEW data to measure employment. The QCEW is a census that incorporates 98 percent of all jobs in the economy. In contrast, the LAUS data is modeled, including some state-level inputs. This use of state-level variables reintroduces some of the endogeneity that the border-county-pair analysis was intended to avoid.[44] The third important difference is the difference in the sample period. As shown in Figure 4, in the baseline BCP sample, there is a downward trend throughout the sample period reducing employment on the treated side and continuing after the treatment differential within the pair is removed. HKMM's use of a long pretreatment window and their early truncation of the sample in 2012 induce a correlation between UI generosity and differential unemployment duration within county pairs. Our symmetric window, by contrast, orthogonalizes the trend and treatment. As shown in Table 3, when we trim a quarter of the county pairs on match quality, we rid the entire sample of these trends (Figure 5), and the sample period no longer substantively affects the estimates. We elaborate on these arguments and discuss the decomposition exercises in greater detail in online Appendix B.[45]

[44] HKMM acknowledge this issue with the LAUS data and take some steps to address it. In particular, they repeat their analysis using a version of the data that strips out some (but not all) of the state-level inputs into the estimated county unemployment rate. Consistent with our analysis above, the resulting estimate using the modified data (0.043) is smaller than their original estimate (0.049). This difference translates to roughly 0.4 percentage points of EPOP, only modestly smaller than the average marginal effect of switching from LAUS to the QCEW (0.67) reported in our online Appendix Table B1.

[45] We also show, in online Appendix Table B2, three examples of full transition paths from HKMM's estimates to our estimates.

B. *HMM Comparison*

Our 2014 event study uses similar variation to HMM but again has very different results. Since HMM presents three main estimation methods, and since the differences between their strategies and ours are more pronounced than our full sample estimates are with HKMM, we do not present a decomposition. Instead, we focus upon one particularly important choice by HMM: the use of LAUS unemployment data.

The HMM estimates are approximately one-third the size of the HKMM estimates; however, they are still large enough to explain the entirety of the 2014 employment boom as resulting from the expiration of the EUC program. In contrast, our estimates are very small in magnitude and statistically insignificant. HMM present three main models: (i) a fixed effects difference-in-difference model with county-specific linear trends and pair-period fixed effects; (ii) a similar model replacing the fixed effects and trends with Bai (2009) interactive effects; and (iii) a model with additional covariates such as the price of oil, aggregate construction employment, reserve balances with the Fed system, and county-specific coefficients on these covariates. They estimate effects comparing 2014 to 2013:IV outcomes.

We first replicate the three HMM models. In all cases, our estimates are within 5 percent of HMM's estimates. Then, we reestimate with two different dependent variables. First, in 2015, the Bureau of Labor Statistics redesigned the LAUS. HMM use the pre-redesign LAUS. We reestimate with our replication of their models using post-redesign, and we also reestimate with the QCEW. We present these estimates in online Appendix Table B3.

Our replication of HMM yields statistically significant results with a 95 percent level of confidence for the first two models and statistical insignificance for the third factor model. In all cases, using the LAUS redesign drops the coefficients by between 78 and 97 percent and yields estimates with t-statistics below 1. The QCEW estimates display greater similarity in magnitude to the postrevision LAUS estimates than either the postrevision LAUS or the QCEW estimates display relative to the pre-redesign LAUS. The QCEW estimates are also all statistically insignificant and between 53 and 88 percent smaller than the corresponding HMM estimates. Thus, the HMM estimates become much smaller and statistically insignificant when the QCEW is used (as with the HKMM estimates), but they also do so when the newest version of the LAUS data is used. We provide a more detailed discussion of the HMM estimates in online Appendix B.

VI. Rationalizing Macro and Micro Effects of UI Extensions

Our estimates represent a "macro" effect of UI extensions on aggregate employment. Most of the literature on the impacts of UI has focused only on the impacts on labor supply behavior. In this section, we compare some of the key "micro" estimates from the literature to our "macro" estimates and provide a discussion of plausible channels that can rationalize the gap between these estimates.

We begin by translating our macro estimates, as well as micro estimates from the literature, into numbers of net jobs created or destroyed. This entails multiplying our

Table 5—Rationalizing Micro and Macro Employment Effects of UI

			ΔE_{MACRO} (in millions)		GAP in ΔE (in millions)	
	β_{MICRO}	ΔE_{MICRO}	Point estimate	Confidence interval	Point estimate	Confidence interval
Rothstein (2011), lower bound	0.1	−0.1	0.5	[−0.9, 1.8]	−0.6	[−1.9, 0.7]
Farber and Valetta (2015)	0.4	−0.5	0.5	[−0.9, 1.8]	−1.0	[−2.3, 0.3]
Rothstein (2011), upper bound	0.5	−0.7	0.5	[−0.9, 1.8]	−1.1	[−2.5, 0.2]
Daly et al. (2012)	0.8	−1.1	0.5	[−0.9, 1.8]	−1.5	[−2.9, −0.2]
Katz and Meyer (1990)	1.3	−1.7	0.5	[−0.9, 1.8]	−2.2	[−3.5, −0.9]
Elsby et al. (2010), upper bound	2.4	−3.2	0.5	[−0.9, 1.8]	−3.7	[−5.0, −2.3]
Johnston and Mas (2018)	4.6	−6.2	0.5	[−0.9, 1.8]	−6.6	[−7.9, −5.3]

Notes: Column 1 displays a range of micro estimates based on other studies, where β_{MICRO} is an estimate of the change in the unemployment rate resulting from only the micro-level effect of a 73-week increase in maximum UI duration. Column 2 displays the corresponding impact on employment (in millions of workers), $\Delta E_{MICRO} = \beta_{MICRO} \times L$, where L is the size of the labor force, expressed in millions. The point estimate in column 3 is the estimated impact on employment (in millions of workers) implied by the results in this paper, $\Delta E_{MACRO} = \beta_{MACRO} \times P$, where P is the population and β_{MACRO} is a direct estimate of the aggregate change in EPOP (from column 2 in panel A of Table 2). The gap in ΔE is calculated as the difference between ΔE_{MACRO} and ΔE_{MICRO}.

estimates (which are in terms of EPOP) by the 15+ population in 2012 (253 million) and the micro estimates (which are in terms of unemployment rates) by the 2012 labor force (134 million). The gap between the macro and the micro estimates of the UI extensions on employment can be written as

$$(8) \qquad \Delta E_{GAP} = \Delta E_{MACRO} - \Delta E_{MICRO} = \left(\beta_{MACRO} \times P + \beta_{MICRO} \times L\right),$$

where β_{MICRO} is a micro estimate from the empirical literature of the impact of raising the UI benefit duration from 26 to 99 weeks on the unemployment rate, L is the size of the labor force (in 2012), β_{MACRO} is an estimate from this paper, and P is the 15+ population in 2012. The resulting ΔE_{MACRO} is the predicted change in national employment resulting from increasing UI benefit duration from 26 to 99 weeks using our estimates, while ΔE_{MICRO} is the analogous predicted employment change using micro estimates from the literature. We then compute ΔE_{GAP} as the unexplained gap between the implied macro employment impact of UI and the implied micro employment impact of UI.

In Table 5, we report computations using 6 estimated micro responses to the impact of increasing UI duration from 26 to 99 weeks in the literature.[46] Five of these micro estimates come from four papers using data from the Great Recession

[46] All of these papers estimate the impact of UI by examining individual employment outcomes. In order to arrive at an estimate of the microelasticity in this type of setting, it is necessary to compare individuals facing similar labor market conditions—either by controlling for macro conditions like the local unemployment rate and labor market tightness or by comparing people within the same labor market. The studies listed in Table 5 vary in the extent to which they control for macro conditions; several of these studies compare individuals facing different labor markets and would thus be more accurately characterized as "mixed" estimates, as discussed in Landais, Michaillat, and Saez (2018). Rothstein (2011), Johnston and Mas (2018), and Katz and Meyer (1990) all provide true micro estimates by making use of specifications comparing individuals within the same labor market. On the other hand, Daly et al. (2012); Elsby, Hobijn, and Şahin (2010); Farber and Valletta (2015); and some of the specifications in Rothstein (2011) produce estimates that are more "mixed."

(Rothstein 2011, Daly et al. 2012, Farber and Valletta 2015, Johnston and Mas 2018). Four of these numbers range between 0.1 and 0.8. Johnston and Mas (2018) is substantially larger in magnitude at 4.6.[47] We also use two estimates from before the Great Recession that come from Katz and Meyer (1990) (1.3) and Elsby, Hobijn, and Şahin (2010) (2.4).[48] We take $\beta_{MACRO} = 0.180$ (from column 2 of Table 2). This suggests a national employment increase of around 0.5 million from the policy. If we use the 95 percent confidence interval, our estimate suggests employment changes ranging between -0.9 million and 1.8 million. In contrast, the implied employment changes based on the micro elasticities range between -6.2 million and -0.1 million; excluding the Johnston and Mas (2018) estimate, the range is -3.2 million to -0.1 million. The employment gap (ΔE_{GAP}) implied between the macro and micro estimates ranges between -0.6 million (when using Rothstein 2011's lower bound) and -6.6 (when using Johnston and Mas 2018). The point estimates from Johnston and Mas (2018); Elsby, Hobijn, and Şahin (2010); Katz and Meyer (1990); and Daly et al. (2012) all imply employment effects outside of our 95 percent confidence interval. In contrast, the estimates from Rothstein (2011) and Farber and Valletta (2015) imply employment losses that fall within our confidence interval. Overall, the evidence broadly suggests that our macro estimate is more positive than the employment losses predicted by the micro estimates, though the lack of sufficient precision warrants caution.

With the precision caveat in mind, we consider explanations for why a macro effect might be more positive than the micro effect. First, since the gap between the macro and the micro estimates is positive, we cannot explain the gap from the vacancy creation effect (Mitman and Rabinovich 2015), as their mechanism is only capable of explaining a more negative macro than micro effect. Landais, Michaillat and Saez (2018), by contrast, do predict a positive ΔE_{GAP}, consistent with our empirical findings. They show that if jobs are rationed during a downturn, then a decrease in labor market search intensity by unemployed individuals due to a more generous UI policy will tend to increase labor market tightness—i.e., the job-finding probability of other unemployed workers. An increase in potential benefit duration reduces the "rat race" between unemployed workers, increases labor market tightness, and implies that $\Delta E_{MACRO} \geq \Delta E_{MICRO}$. In their model, this "wedge"

[47] As we noted in the introduction, Johnston and Mas (2018) provide a case study of Missouri, where there was a sudden reduction in benefits, and find a much larger micro-level response than most of the literature does. Besides providing labor supply based estimates, they also provide synthetic control and difference-in-difference estimates for aggregate employment effects from the benefit reduction. These macro estimates are similarly sized to their micro estimates and much larger than the macro effects that we find in this paper. Therefore, the size of the estimates from Johnston and Mas (2018) seem less about the micro versus macro effects and more about the Missouri case study. Nonetheless, here we include the implied β_{MICRO} estimates from the Johnston and Mas (2018) study since those are specifically based on the labor supply response to the policy change.

[48] We translate β_{MICRO} for the Katz and Meyer (1990) estimates using the approach in Mazumder (2011). Kroft and Notowidigdo (2016) show that responsiveness to UI declines during recessions and thus the estimates from both Katz and Meyer (1990) and Elsby, Hobijn, and Şahin (2010) are likely larger than UI impacts upon labor supply during the Great Recession. We additionally note that the Katz and Meyer (1990) estimates may be inappropriate for studying the Great Recession, because a substantial part of their estimate reflects recalls from layoffs—which are much less common today (see Rothstein 2011 for a discussion on this). We also note that the estimates in Landais, Michaillat, and Saez (2018) using the same data and a regression kink design yield similar magnitudes as Katz and Meyer (1990).

between the micro and macro elasticities depends on how tightness responds to UI. Our findings are broadly consistent with a positive wedge.

At the same time, this "rat race" channel in isolation cannot explain a positive macro effect of UI, $\Delta E_{MACRO} > 0$, as suggested by our point estimates (though our confidence interval contains zero). Instead, a positive employment effect could be explained by a Keynesian demand channel. UI puts cash in the hands of unemployed individuals whose earnings in the absence of UI payments are likely to be well below their permanent incomes. These individuals are likely to be liquidity constrained, and thus a dollar of UI expenditures is highly likely to be consumed (Ganong and Noel 2019). In the specific context of the the EB and EUC programs, the extension of benefits led to net transfers to local areas where benefit durations were increasing—and existing research suggests that these likely had a stimulative effect on employment. For example, reviewing the literature using ARRA stimulus during the Great Recession, Chodorow-Reich (2019) estimates a cross-sectional multiplier of around 1.8. To get a sense of the implied stimulus effect from increasing potential benefit duration from 26 to 99 weeks, we multiply the hypothetical increase in federal expenditures of \$74.4 billion[49] by 1.8 and divide by the ratio of output to employment ($Y/E = \$108,000$).[50] These back-of-the-envelope calculations suggest that the stimulus effect of the policy would create an additional 1.2 million jobs.

These implied job gains from a pure aggregate demand effect are larger than the 0.5 million jobs suggested by our upper bound macro estimate, though they are within our confidence interval. At the same time, the implied jobs gains are comparable to the gap between our estimate and estimates from Farber and Valletta (2015), Rothstein (2011), and Daly et al. (2012). These calculations provide a rough sense that while some stimulus effect is needed to rationalize a positive macro effect, the size of the macro effect is likely smaller than would be expected if we only considered the aggregate demand channel. To actually disentangle the labor supply effect, the rat race effect, the vacancy creation effect (as in Mitman and Rabinovich 2015), and the aggregate demand effects requires a full-fledged model that incorporates these elements. One such example comes from Kekre (2019), who calibrates a search-and-matching model with nominal rigidities. He finds that the UI extensions during the Great Recession had a small positive impact upon the employment-to-population ratio. His results are qualitatively and quantitatively consistent with our empirical findings. Going forward, empirically separating these various channels represents an important area for future research.

[49] National EB and EUC transfer payments between November 2008 and December 2013 averaged \$49.3 billion annually, and during this time period, the average number of weeks of UI available was 74.4. In order to obtain an estimate of UI expenditures corresponding to an increase from 26 to 99 weeks, we scale the actual expenditure by $(99-26)/(74.4-26)$ ($\Delta B = \$49.3 \times 10^9 \times (73/48.4) = \$74.4 \times 10^9$). We obtain the data for payments made through the EB and EUC programs from http://oui.doleta.gov/unemploy/euc.asp.

[50] GDP per worker data from 2012 are from the World Bank (http://data.worldbank.org/indicator/SL.GDP.PCAP.EM.KD?locations=US). Our estimates closely follow the approach in Chodorow-Reich (2019) and implicitly assume that jobs created from the fiscal stimulus have mean productivity; Chodorow-Reich (2019) provides evidence supporting the validity of this approximation. Similarly, Nakamura and Steinsson (2014) report both output and employment multipliers using defense spending shocks, and the magnitudes of both are are consistent with this approximation.

VII. Conclusion

Despite a large literature that has evaluated the labor supply effects of UI, the overall impact of the policy on aggregate employment is a relatively new and understudied area of research. Yet, it is an important question from a public policy perspective. If there are sizable negative effects of UI upon employment via labor supply but these are counteracted by positive aggregate demand effects, then the overall employment effects can be more positive than what is implied by the labor supply estimates—making the policy more effective. Conversely, if the labor supply effects are small but higher reservation wages fuel lower hiring and hence a higher unemployment rate, then the policy can be less attractive than it may initially appear from micro evidence alone.

In this paper, we add to the small but growing literature on the impact of UI on overall employment. We utilize variation across counties that straddle state borders where the states differ in their UI duration during the Great Recession. We find that this strategy substantially reduces likely bias from endogeneity that would plague a two-way fixed effects model assuming parallel trends across counties (or states) receiving differential treatment. To account for remaining endogeneity, we utilize a variety of strategies, including refining our sample and focusing on variation driven by the national policy changes created by the introduction of differential EUC across states in 2008 as well as the expiration of the EUC program at the end of 2013.

Whether we use all policy variation or we use variation induced solely by national-level policy changes, most of our estimates are quite small in magnitude. Our full sample results using a refined border-county-pair design suggest that the employment to population ratio rose by a statistically insignificant 0.180 percentage points due to the 73-week increase in benefits. The event study results that use the national policy variation from 2008 expansion and 2014 expiration of EUC suggest the EPOP ratio increased by 0.253. While the 95 percent confidence intervals for the full sample estimate rules out change in EPOP more negative than -0.345, the confidence bounds for the event study rule out changes more negative than -1.021. Finally, our dynamic specifications do not indicate any policy anticipation effects.

Overall, our findings are similar to recent estimates by Chodorow-Reich, Coglianese, and Karabarbounis (2019), who use policy variation that is quite different from what we use in this paper, as well as estimates by Dieterle, Bartalotti, and Brummet (2020), who use a regression discontinuity design. At the same time, our estimates and conclusions are quite different from those reached by HMM and HKMM, even though they also use a border-county-pair-based strategy. The differences are in large part due to three main choices: their use of bias-inducing auxiliary parametric assumptions that we do not find to be warranted by the data, their use of an incomplete portion of the treatment window, and their use of (model-based) LAUS data.

The small macro employment effects of UI found in this paper are consistent with small negative effects on labor supply typically (though not always) found in the existing literature together with an impact on labor market tightness and a moderately sized, positive effect on aggregate demand in the local economy. Future research should better assess the relative contributions of these two macro channels.

Nonetheless, our results suggest that the overall employment impact of the sizable UI extensions during the Great Recession was likely modest. At worst, they led to a small reduction in aggregate employment, and at best, they slightly boosted employment in the local economy.

REFERENCES

Amaral, Pedro, and Jessica Ice. 2014. "Reassessing the Effects of Extending Unemployment Insurance Benefits." Federal Reserve Bank of Cleveland. https://www.clevelandfed.org/en/newsroom-and-events/publications/economic-commentary/2014-economic-commentaries/ec-201423-reassessing-the-effects-of-extending-unemployment-insurance-benefits.aspx.

Bai, Jushan. 2009. "Panel Data Models with Interactive Fixed Effects." *Econometrica* 77 (4): 1229–79.

Boone, Christopher, Arindrajit Dube, Lucas Goodman, and Ethan Kaplan. 2021. "Replication data for: Unemployment Insurance Generosity and Aggregate Employment." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.38886/E118525V1.

Borusyak, Kirill, and Xavier Jaravel. 2017. "Revisiting Event Study Designs, with an Application to the Estimation of the Marginal Propensity to Consume." Unpublished.

Bureau of Labor Statistics. 1976–2015. "Local Area Unemployment Statistics [2005 Redesign]." FRED API. https://fred.stlouisfed.org/docs/api/fred/ (accessed October 22, 2015).

Chodorow-Reich, Gabriel. 2019. "Geographic Cross-Sectional Fiscal Spending Multipliers: What Have We Learned?" *American Economic Journal: Economic Policy* 11 (2): 1–34.

Chodorow-Reich, Gabriel, John Coglianese, and Loukas Karabarbounis. 2019. "The Macro Effects of Unemployment Benefit Extensions: A Measurement Error Approach." *Quarterly Journal of Economics* 134 (1): 227–79.

Daly, Mary C., Bart Hobijn, Ayşegül Şahin, and Robert G. Valletta. 2012. "A Search and Matching Approach to Labor Markets: Did the Natural Rate of Unemployment Rise?" *Journal of Economic Perspectives* 26 (3): 3–26.

Dieterle, Steven, Otávio Bartalotti, and Quentin Brummet. 2020. "Revisiting the Effects of Unemployment Insurance Extensions on Unemployment: A Measurement-Error-Corrected Regression Discontinuity Approach." *American Economic Journal: Economic Policy* 12 (2): 84–114.

Dube, Arindrajit, T. William Lester, and Michael Reich. 2010. "Minimum Wage Effects across State Borders: Estimates Using Contiguous Counties." *Review of Economics and Statistics* 92 (4): 945–64.

Dube, Arindrajit, T. William Lester, and Michael Reich. 2016. "Minimum Wage Shocks, Employment Flows, and Labor Market Frictions." *Journal of Labor Economics* 34 (3): 663–704.

Elsby, Michael W.L., Bart Hobijn, and Ayşegül Şahin. 2010. "The Labor Market in the Great Recession." *Brookings Papers on Economic Activity* 40 (1): 1–48.

Employment and Training Administration. 2003–2015a. "Extended Benefits Trigger Notice." US Department of Labor. https://oui.doleta.gov/unemploy/claims_arch.asp.

Employment and Training Administration. 1990–2015b. "Significant Provisions of State Unemployment Insurance Laws." US Department of Labor. https://oui.doleta.gov/unemploy/statelaws.asp.

Employment and Training Administration. 2008–2013. "Emergency Unemployment Compensation Trigger Notice." US Department of Labor. https://oui.doleta.gov/unemploy/claims_arch.asp.

Farber, Henry S., Jesse Rothstein, and Robert G. Valletta. 2015. "The Effect of Extended Unemployment Insurance Benefits: Evidence from the 2012–2013 Phase-Out." *American Economic Review* 105 (5): 171–76.

Farber, Henry S., and Robert G. Valletta. 2015. "Do Extended Unemployment Benefits Lengthen Unemployment Spells? Evidence from Recent Cycles in the U.S. Labor Market." *Journal of Human Resources* 50 (4): 873–909.

Ganong, Peter, and Pascal Noel. 2019. "Consumer Spending during Unemployment: Positive and Normative Implications." *American Economic Review* 109 (7): 2383–2424.

Hagedorn, Marcus, Fatih Karahan, Iourii Manovskii, and Kurt Mitman. 2019. "Unemployment Benefits and Unemployment in the Great Recession: The Role of Equilibrium Effects." Federal Reserve Bank of New York Staff Report 646.

Hagedorn, Marcus, Iourii Manovskii, and Kurt Mitman. 2016. "Interpreting Recent Quasi-experimental Evidence on the Effects of Unemployment Benefit Extensions." NBER Working Paper 22280.

Hall, Robert E. 2013. "Some Observations on Hagedorn, Karahan, Manovskii, and Mitman,'Unemployment Benefits and Unemployment in the Great Recession: The Role of Macro Effects.'" https://www.semanticscholar.org/paper/Some-Observations-on-Hagedorn-%2C-Karahan-%2C-Manovskii-Hall-Hoover/8c59c2ac40790204ca5f6321423a6aba7b70f161.

Johnston, Andrew C., and Alexandre Mas. 2018. "Potential Unemployment Insurance Duration and Labor Supply: The Individual and Market-Level Response to a Benefit Cut." *Journal of Political Economy* 126 (6): 2480–2522.

Katz, Lawrence F., and Bruce D. Meyer. 1990. "The Impact of the Potential Duration of Unemployment Benefits on the Duration of Unemployment." *Journal of Public Economics* 41 (1): 45–72.

Kekre, Rohan. 2019. "Unemployment Insurance in Macroeconomic Stabilization." https://scholar.harvard.edu/rkekre/publications/unemployment-insurance-macroeconomic-stabilization-0.

Kroft, Kory, and Matthew J. Notowidigdo. 2016. "Should Unemployment Insurance Vary with the Unemployment Rate? Theory and Evidence." *Review of Economic Studies* 83 (3): 1092–1124.

Lalive, Rafael, Camille Landais, and Josef Zweimüller. 2015. "Market Externalities of Large Unemployment Insurance Extension Programs." *American Economic Review* 105 (12): 3564–96.

Landais, Camille, Pascal Michaillat, and Emmanuel Saez. 2018. "A Macroeconomic Approach to Optimal Unemployment Insurance: Theory." *American Economic Journal: Economic Policy* 10 (2): 152–81.

Marinescu, Ioana. 2017. "The General Equilibrium Impacts of Unemployment Insurance: Evidence from a Large Online Job Board." *Journal of Public Economics* 150: 14–29.

Mazumder, Bhash. 2011. "How Did Unemployment Insurance Extensions Affect the Unemployment Rate in 2008–10?" Federal Reserve Bank of Chicago Fed Letter 285.

Michaillat, Pascal. 2012. "Do Matching Frictions Explain Unemployment? Not in Bad Times." *American Economic Review* 102 (4): 1721–50.

Mitman, Kurt, and Stanislav Rabinovich. 2015. "Optimal Unemployment Insurance in an Equilibrium Business-Cycle Model." *Journal of Monetary Economics* 71: 99–118.

Nakamura, Emi, and Jón Steinsson. 2014. "Fiscal Stimulus in a Monetary Union: Evidence from US Regions." *American Economic Review* 104 (3): 753–92.

Rothstein, Jesse. 2011. "Unemployment Insurance and Job Search in the Great Recession." *Brookings Papers on Economic Activity* 41 (2): 143–214.

Sun, Liyang, and Sarah Abraham. Forthcoming. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." *Journal of Econometrics.*

US Bureau of Labor Statistics. 1990–2015. "Local Area Unemployment Statistics." US Bureau of Labor Statistics. https://www.bls.gov/lau/data.htm (accessed November 10, 2016).

US Bureau of Labor Statistics. 1998–2015. "Quarterly Census of Employment and Wages." US Bureau of Labor Statistics. https://www.bls.gov/cew/downloadable-data-files (accessed November 10, 2016).

US Census Bureau. 2016a. "Population and Housing Unit Estimates." US Census Bureau. https://www.census.gov/programs-surveys/popest.html (accessed November 22, 2016).

US Census Bureau. 2016b. "Quarterly Workforce Indicators." US Census Bureau. https://lehd.ces.census.gov/data/ (accessed November 11, 2016).