Unemployment Insurance Generosity and Aggregate Employment

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

June 14, 2020

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.

*Boone: Cornell University, 530 Statler Hall, Ithaca, NY 14853 (email: cb238@cornell.edu); Dube: University of Massachusetts Amherst, 1030 Thompson Hall, University of Massachusetts Amherst, 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). We thank Gabriel Chodorow-Reich, Thomas Hegland, Ioana Marinescu, Jesse Rothstein, and the editor 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.

During the Great Recession, existing law and new acts of Congress led to the most dramatic expansion in the generosity of unemployment insurance (UI) benefits in U.S. 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 (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 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 et al. 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 U.S. (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

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

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 as well as state unemployment rates. We show that the BCP strategy substantially reduces the endogeneity problem, mitigating negative pre-existing 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

[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 which 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.

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 pre-sample 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 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% 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% 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

[3] HKMM find that the expansion of UI during the Great Recession from 26 to 99 weeks increased the unemployment rate by 80%, 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

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 which additionally includes the phase-out 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 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 pre-trends 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-discontinuities 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

the years after the trough of the Great Recession. HMM study the 2014 expiration of EUC and find that that expiration was responsible for the creation of approximately two million jobs. This effect would translate into a 1.1% 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 data set.

[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.

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 unemployment insurance 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 which is comparable to the gaps we find relative to most of the micro estimates from the Great Recession in the U.S. 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 unemployment insurance 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 unemployment insurance receipt increased from 26 weeks (in all but two states) to as much as 99 weeks.[5] The path of this expansion and subsequent phase-out 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 to 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 pre-existing 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, as well as 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.

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

[5] Before the Great Recession, Massachusetts and Montana both had more than 26 weeks of unemployment insurance; Montana had a maximum of 28 and Massachussetts had a maximum of 30.

[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.

across neighboring counties in the numbers of weeks of available unemployment insurance, where the reported difference is between "high" and "low" benefit duration counties, defined by comparing the average duration in the treatment period (2008m11-2013m12) versus the surrounding 24 months (2007m11-2008m10 and 2014m1-2014m12) when these differences were typically zero or very small.

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 zero 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. **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 2016*b*). 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% of jobs are covered by unemployment insurance, 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 data set 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.

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 (U.S. Census Bureau 2016*a*). 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

[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 unemployment insurance system, such as railroad workers and workers at religiously-affiliated schools.

[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.

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 which are issued biannually (Employment and Training Administration 1990–2015*b*). 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 2003–2015*a*, 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 that 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 pre-lapse level.

We also use a list of all contiguous county pairs that straddle state borders; this data comes 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 2016*a*). We additionally obtain a vintage series of county unemployment rates and employment (prior to the March 2015 redesign) via FRED (Bureau of Labor Statistics 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.

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

[9] The time series of aggregate EPOP is displayed in **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.

to contiguous county pairs which 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 are 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 pre-treatment covariate balance within border county pairs, and the absence of pre-existing 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 non-treatment months was 30 weeks, 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 which straddle state borders during the treatment period. **Table 1** shows that the "high" treatment and "low" treatment counties were quite similar. Pre-existing characteristics are relatively balanced within pairs between the counties that received high versus low average treatments. Only one of twelve covariates (share with a college degree) is statistically significantly different at conventional levels.[11]

We also present aggregate graphical evidence on pre-existing trends and treatment effects. In particular, we regress EPOP on a set of $treat_c \times \mathbb{1}\{t = s\}$ variables, where $\mathbb{1}\{t = s\}$ 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]

[10]This "non-treatment" 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 **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 **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.

[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

$$E_{cpt} = \sum_{s=\tau_A}^{\tau_B} \beta_s treat_c \mathbb{1}\{t = s\} + \lambda_c + \nu_{pt} + \epsilon_{cpt} \tag{1}$$

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) pre-existing 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. 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.

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 (1) took the spatial difference of the dependent variable and main independent variable across each county pair p at each time t, and (2) replaced county fixed effects by pair fixed effects.

[14] **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 county border pairs is largely immaterial.

[15] **Appendix Table C3** quantifies the magnitude of these pre-existing 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 pre-existing 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.

Trimming on pre-treatment match quality

While the evidence in **Figure 4** shows that the BCP design mostly eliminates the pre-existing 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 pre-sample EPOP. Restricting the analysis to pairs that are better-matched based on pre-treatment 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: 1) pre-treatment mean squared error (MSE) in employment between counties, 2) absolute differences in pre-treatment linear time trend in employment between counties, and (3) differences in pre-treatment demographic and economic covariates between counties.[16] For each of these three criteria, we estimate a "match quality" variable in the "training sample" of 2003m11-2007m10. Next, we trim the worst 25% of the pairs based on each criteria. Finally, we use a "test sample" of the 2007m11-2008m10 period prior to treatment to assess the out-of-sample MSE in employment. We choose the trimming criteria which has the minimal out-of-sample MSE. We find that pre-treatment MSE-trimming performs the best: as compared to the baseline BCP sample, it reduces out-of-sample MSE by 65%. In contrast, trimming on pre-treatment linear trends reduces out-of-sample mean MSE by 44% and covariate trimming results in a 25% reduction. Therefore, throughout the paper, we show estimates based on the baseline BCP-FE sample as well as the Pre-Treatment-trimmed (henceforth PT-trimmed) sample which drops the county pairs in the top quartile of pre-treatment 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 pre-existing 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 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

[16] In particular, we construct a Mahalanobis measure incorporating within-pair differences in share white, share Hispanic, share with a bachelors degree or more, log median per capita income, share urban, number of mortgage originations in 2007, log population, 2006 EPOP, and share of employment in the goods industry.

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 which exploits only the variation induced by national-level policy changes to the EUC program in 2008 and 2014.

IV.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 which 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 top panel 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 pre-treatment 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% confidence

[17] 32 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 data set and thus potentially differently in some pairs compared to others.

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 pre-existing 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 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 twelve months prior to treatment, i.e., between $\tau = -12$ and -1, there is no statistically or economically significant change in employment. The 12^{th} 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

[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 **Appendix Table C4**.

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.

IV.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 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 2018 and Sun and Abraham 2020) 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 county border 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 provides a cleaner test of pre-existing 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%. This led to an increase in UI benefit durations which varied

[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.

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 which 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 six 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% saw a bigger increase in benefit duration than 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 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:

[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.

$$z_{ct} = \begin{cases} D_c^{08} & \text{Nov. 2007 - Oct. 2008} \\ D_c^{08} + \delta_c^{08} & \text{Nov. 2008 - Oct. 2009} \\ D_c^{13} & \text{Jan. 2013 - Dec. 2013} \\ D_c^{13} - \delta_c^{13} & \text{Jan. 2014 - Dec. 2014} \end{cases}$$

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 $\left(\delta_c^{13}\right)$, 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 $\left(D_c^{08}\right)$ and just after the introduction of the new EUC program. Therefore, the jump in z_{ct} that occurs in November 2008 $\left(\delta_c^{08}\right)$ 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 panel 2 of **Table 2**, we report our pooled results using both the 2008 introduction (i.e., a positive treatment) and the 2014 expiration of the EUC (i.e., a negative treatment). 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 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 which 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.

[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.

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 2007m11 to 2009m10 period are reported in the third panel of **Table 2** and the 2013m1-2014m12 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 (2/3 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% 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% 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]

$$\Delta E_{cpt} = \sum_{\tau=-12}^{11} \beta_\tau \delta_{ct} \mathbb{1}\{eventdate_t = \tau\} + \nu_{pt} + \eta_{cpt} \tag{6}$$

$$\Delta D_{ct} = \sum_{\tau=-12}^{11} \beta_{z\tau} \delta_{ct} \mathbb{1}\{eventdate_t = \tau\} + \gamma_{pt} + \epsilon_{cpt} \tag{7}$$

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. The top figure displays the pooled estimates while the bottom figures show results for the 2008 and 2014 events separately. The corresponding results for the PT-trimmed sample are displayed in **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 both in 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 national-level policy changes. Furthermore, the results are visually similar both in the baseline BCP-FE

[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 10 weeks because the policy changes in question did not occur precisely at the end of a calendar month.

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% of their immediate pre-expiration 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 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. And 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.

IV.C Robustness of estimates

IV.C.1 Choice of sample period

Table 3 shows results from the full sample specification for alternative sample periods beginning in 2007m11, 2006m11, 2005m11, and 2004m11.[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** is statistically significantly different from zero at conventional levels.

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 to the sample starting in November 2004 is non-trivial; it represents a differential impact of 0.76 percentage points of EPOP from a 73-week

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

increase in UI duration. This is almost twice the baseline estimate starting in 2007m11. 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 2007m11-2014m12 sample frame—with twelve 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 are much smaller in PT-trimmed sample, though we cannot say conclusively whether this primarily reflects a reduction in noise or in residual endogeneity.

IV.C.2 Trimming on pre-treatment mean-squared error

The refined BCP strategy trims the pairs with the worst matches—25% of the sample with the biggest out-of-sample mean squared prediction error in pre-treatment EPOP. In **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 7 rows. The rows are, respectively: no trimming, 10% trimming, 20% trimming, 25% trimming, 30% trimming, 40% trimming, and trimming at the median of the difference in pre-treatment MSE. The 25% 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% trim. It is minimized at a 50% trim for the pooled event study sample, 40% for the 2008 sample and 10% for the 2014 sample. Thus, our choice of a 25% 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 variation. It does not seem to systematically change the magnitude of the estimate in a positive or in a negative direction. The reduction in the standard errors is often up to 50% from the baseline sample. The one exception is the 2014 estimate where the

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

maximum reduction across trimming thresholds is approximately 20%.

IV.C.3 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 both for the baseline BCP sample as well as 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 2004m11-2008m10 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 pre-treatment MSE estimated over the 2004m11-2007m10 period (instead of 2004m11-2008m10). Because the temporary lapses in EUC extensions in the absence of Congressional re-authorization (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 data set, the Quarterly Workforce Indicators (QWI, U.S. Census Bureau 2016*b*). 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 states.[34] In the sixteenth row, to demonstrate that our controls are well matched to our

[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 $((\frac{99}{26})^{.005} - 1) \times 42 = 0.282$ 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.

[34] Recall that North Carolina lost access to EUC at the end of June 2013. This was a full 6 months before the other states lost

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 km 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% 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% 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 **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% 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 2 year sample period. In addition, they are imprecise because of the large variation in EPOP during the onset of the Great Recession.

IV.C.4 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 by computing the population-weighted average distance to the border for every county along each state border segment. The addition of spatial

access to EUC benefits, which means that North Carolina gets treated half way 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 (Row 14) include North Carolina in the 2014 portion of the pooled event study specification. 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 will discuss row 17 below, in Section IV.C.4.

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). **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-differences (or difference in discontinuities) 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 **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 and 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 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. This is to be expected since, effectively, the Dieterle, Bartalotti and Brummet (2020) estimator is estimating off of a substantially smaller set of counties which are particularly geographically close. There are non-border ("hinterland") counties which are close to the border, and we can add Hinterland County Pairs (HCP) to our estimation to gain precision when using the regression discontinuity design. Intuitively,

[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) which 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 **Appendix Table C8**.

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 **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 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% confidence level. We explicitly test for and cannot reject at conventional levels that the two sets of estimates (BCP and HCP) are the same.

Additionally, in **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 inter-centroid 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.

IV.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-

[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).

level changes in benefit duration that took place during the Great Recession. **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% 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% 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 **Appendix Figure C6,** we show that similar 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, ten weeks after the event, 70% of the original gap in maximum benefit duration between the two sides of the border remained in place. Even 52 weeks after the event, on average, more than 50% 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 policy persistence was widespread across counties. In panel A of **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% of the county pairs; by 40 weeks, about 15% 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% 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

[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 zero 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 $\tau = 0, 1, ..., 51$) would be equal to $D_{A,t+\tau} - D_{B,,t+\tau} - 6$, since the pre-shock difference was 6 weeks. Therefore, the regression coefficients trace out the share of the original shock that remains after τ weeks.

specification, Chodorow-Reich, Coglianese and Karabarbounis (2019) uses treatment events whose half life is roughly 8 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]

V.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 non-consequential choices are: (1.) we use levels of variables whereas HKMM use logs, (2.) we control for county fixed effects and time fixed effects in our model whereas HKMM use interactive fixed effects, and (3.) 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: (4.) HKMM quasi-forward difference their dependent variable whereas we do not, (5.) we use employment from the QCEW to construct our main dependent variable whereas HKMM use unemployment from LAUS, and (6.) we use a symmetric 1-year window surrounding the time period of differential UI expansion (2007m11-2014m12) whereas HKMM use the time period 2005m1-2012m12.

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 **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

[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.

[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.

five other assumptions.[42] In other words, we separately estimate the average marginal impact of moving from HKMM's to our estimates across all configuration 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% is due to the use of quasi-forward differencing, 30% is due to alignment of the sample period, and 22% is due to the usage of LAUS data as opposed to QCEW data (column 3 of **Appendix Table B1**). The marginal effects for the first two of these three changes are statistically significant at the 99% confidence level, while the third is significant at the 90% confidence level. The average marginal effect of the other three changes are substantially smaller and with t-statistics well below 1. The use of interactive as opposed to linear fixed effects account for only 6% of the difference and the use of logs as opposed to levels as well as the alignment of the county sample each explain less than 0.3% 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 **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 which incorporates 98% 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

[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 $\frac{6!}{2}$ 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 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 from our estimate are larger than the averages along the transition path, reflecting that the various differences between the two estimates are substitutes.

[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 which 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 **Appendix Table B1**.

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 pre-treatment window and their early truncation of the sample in 2012 induces 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]

V.B HMM Comparison

Our 2014 event study uses similar variation to HMM but again with 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 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 1/3 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: (1.) a fixed effects difference-in-differences model with county-specific linear trends and pair-period fixed effects, (2.) a similar model replacing the fixed effects and trends with Bai interactive effects, and (3.) a model with additional covariates such as the price of oil, aggregate construction employment, and reserve balances with the Fed system; they estimate county-specific covariate coefficients with this third model. They estimate effects comparing 2014 to 2013Q4 outcomes.

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

Our replication of HMM yields statistically significant results with a 95% levels 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% and yields estimates with t-statistics below 1. The QCEW estimates display greater similarity in magnitude to the post-revision LAUS estimates than either the post-revision LAUS or the QCEW estimates display relative to the pre-redesign LAUS. The QCEW estimates are

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

also all statistically insignificant and between 53% and 88% 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 also 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 which 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 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:

$$\Delta E_{GAP} = \Delta E_{MACRO} - \Delta E_{MICRO} = (\beta_{MACRO} \times P + \beta_{MICRO} \times L) \quad (8)$$

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 estimated using data from the Great Recession (Rothstein 2011; Daly et al. 2012; Farber and Valletta

[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 provides 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".

2015; Johnston and Mas 2018). Four of these numbers range between 0.1 to 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 which 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% confidence interval, our estimate suggests employment changes ranging between -0.8 million and 1.9 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 -0.7 million (when using Rothstein's lower bound) and -6.7 (when using Johnston and Mas). 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% 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, does 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" between the micro and macro elasticities depends on how tightness responds to UI. Our findings are broadly consistent with a positive 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. 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 as their micro estimates, and are 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 than about the Missouri case study. Nonetheless, here we include the implied β_{MICRO} estimates from 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).

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 $\left(\frac{Y}{E} = \$108,000\right)$.[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, 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 $\frac{99-26}{74.4-26}$ ($\Delta B = \$49.3 \times 10^9 \times \frac{73}{48.4} = \$74.4 \times 10^9$. 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 is 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 assumes 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 unemployment insurance, 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, 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 fuels lower hiring and hence a higher unemployment rate, 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 which 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 whether 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 suggests the EPOP ratio increased by 0.253. While the 95% confidence intervals for the full sample estimate rules out change in EPOP more negative than -0.345, the confidence bounds for the event studyrule 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 which 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."

Bai, Jushan. 2009. "Panel data models with interactive fixed effects." *Econometrica*, 77(4): 1229–1279.

Borusyak, Kirill, and Xavier Jaravel. 2018. "Revisiting Event Study Designs, with an Application to the Estimation of the Marginal Propensity to Consume."

Bureau of Labor Statistics. 2015. "Local Area Unemployment Statistics (1976–2015) [2005 Redesign]." retrieved from FRED API, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/docs/api/fred/ (Accessed October 22, 2015).

Bureau of Labor Statistics. 2016*a*. "Local Area Unemployment Statistics (1990–2015) [2015 Redesign]." U.S. Department of Labor. https://www.bls.gov/lau/data.htm (Accessed November 10, 2016).

Bureau of Labor Statistics. 2016*b*. "Quarterly Census of Employment and Wages (1998–2015 NAICS-Based Data Files)." U.S. Department of Labor. https://www.bls.gov/cew/downloadable-data-files.htm (Accessed November 10, 2016).

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." *The Quarterly Journal of Economics*, 134(1): 227–279.

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." *The Review of Economics and Statistics*, 92(4): 945–964.

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*, 1–48.

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

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

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

Farber, Henry S., and Robert G. Valletta. 2015. "Do extended unemployment benefits lengthen unemployment spells? Evidence from recent cycles in the US labor market." *Journal of Human Resources*, 50(4): 873–909.

Farber, Henry S., Jesse Rothstein, and Robert G. Valletta. 2015. "The Effect of Extended Unemployment Insurance Benefits: Evidence from the 2012–2013 Phase-Out." *The American Economic Review*, 105(5): 171–176.

Ganong, Peter, and Pascal Noel. 2019. "Consumer spending during unemployment: Positive and normative implications." *American Economic Review*, 109(7): 2383–2424.

Gomez, Matthieu. 2015. "REGIFE: Stata module to estimate linear models with interactive fixed effects." *Statistical Software Components, Boston College Department of Economics*.

Hagedorn, Marcus, Fatih Karahan, Iourii Manovskii, and Kurt Mitman. 2015. "Unemployment Benefits and Unemployment in the Great Recession: The Role of Macro Effects."

Hagedorn, Marcus, Fatih Karahan, Iourii Manovskii, and Kurt Mitman. 2019. "Unemployment Benefits and Unemployment in the Great Recession: The Role of Equilibrium Effects."

Hagedorn, Marcus, Iourii Manovskii, and Kurt Mitman. 2016. "Interpreting Recent Quasi-Experimental Evidence on the Effects of Unemployment Benefit Extensions." National Bureau of Economic Research, Inc NBER Working Paper 22280.

Hall, Robert E. 2013. "Some Observations on Hagedorn, Karahan, Manovskii, and Mitman,'Unemployment Benefits and Unemployment in the Great Reces sion: The Role of Macro Effects'."

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."

Kroft, Kory, and Matthew J Notowidigdo. 2016. "Should unemployment insurance vary with the unemployment rate? Theory and evidence." *The Review of Economic Studies*, 83(3): 1092–1124.

Lalive, Rafael, Camille Landais, and Josef Zweimüller. 2015. "Market externalities of large unemployment insurance extension programs." *The American Economic Review*, 105(12): 3564–3596.

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, Bhashkar. 2011. "How did unemployment insurance extensions affect the unemployment rate in 2008–10?" *Chicago Fed Letter*, , (Apr).

Michaillat, Pascal. 2012. "Do matching frictions explain unemployment? Not in bad times." *The American Economic Review*, 102(4): 1721–1750.

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 Jon Steinsson. 2014. "Fiscal stimulus in a monetary union: Evidence from US regions." *The American Economic Review*, 104(3): 753–792.

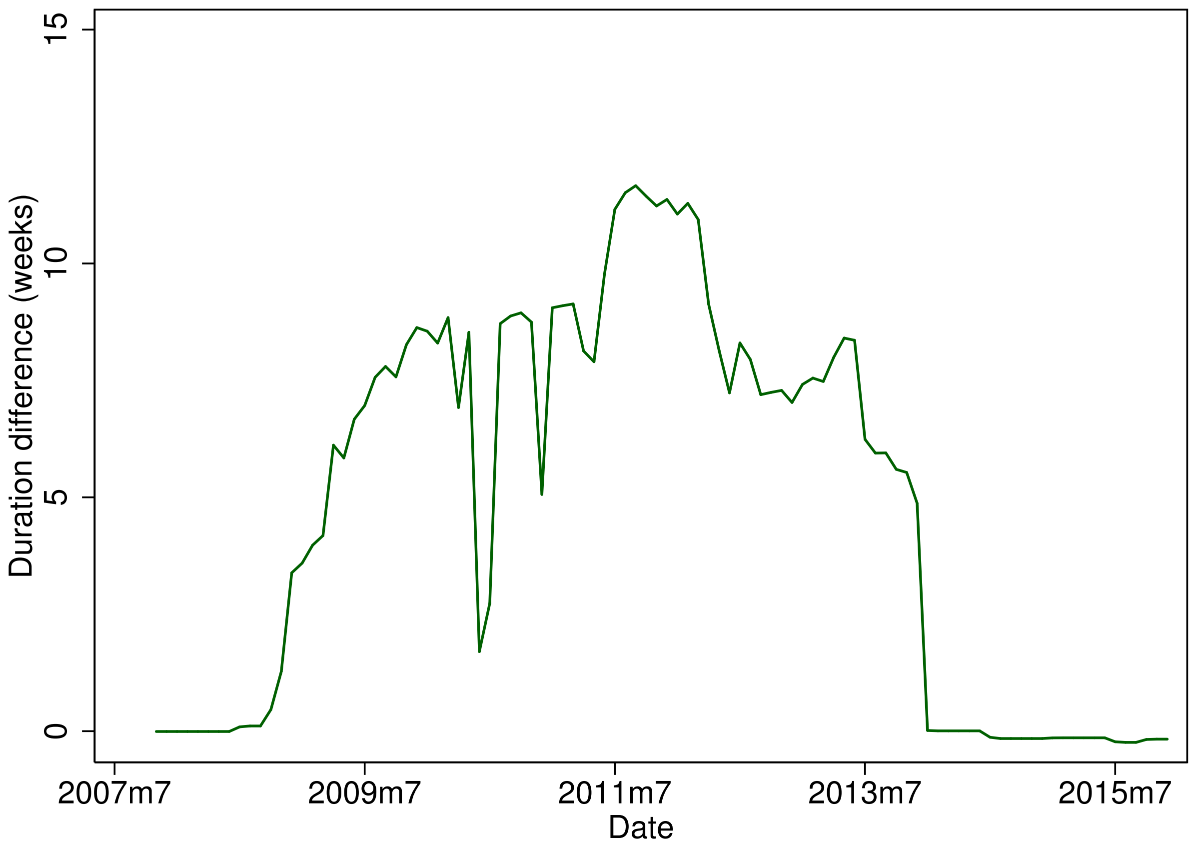
Rothstein, Jesse. 2011. "Unemployment insurance and job search in the Great Recession." *Brookings Papers on Economic Activity*, 143–214.

Sun, Liyang, and Sarah Abraham. 2020. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects."

U.S. Census Bureau. 2016*a*. "Population and Housing Unit Estimates [computer file]." Washington, DC: U.S. Census Bureau, Population Estimates Program [distributor]. https://www.census.gov/programs-surveys/popest.html (Accessed November 22, 2016).

U.S. Census Bureau. 2016*b*. "Quarterly Workforce Indicators (1998-2016) [computer file]." Washington, DC: U.S. Census Bureau, Longitudinal-Employer Household Dynamics Program [distributor]. https://lehd.ces.census.gov/data/ (Accessed November 11, 2016). R2016Q3 [version].

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 in 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 2008m11-2013m12 and average duration from 2007m11-2008m10 and 2014m1-2014m12. The counties in the 30 pairs where this difference is identical are assigned arbitrarily to the “high” and “low” sets.

Figure 2: Reduction in UI benefit duration from the December 2013 expiration of EUC

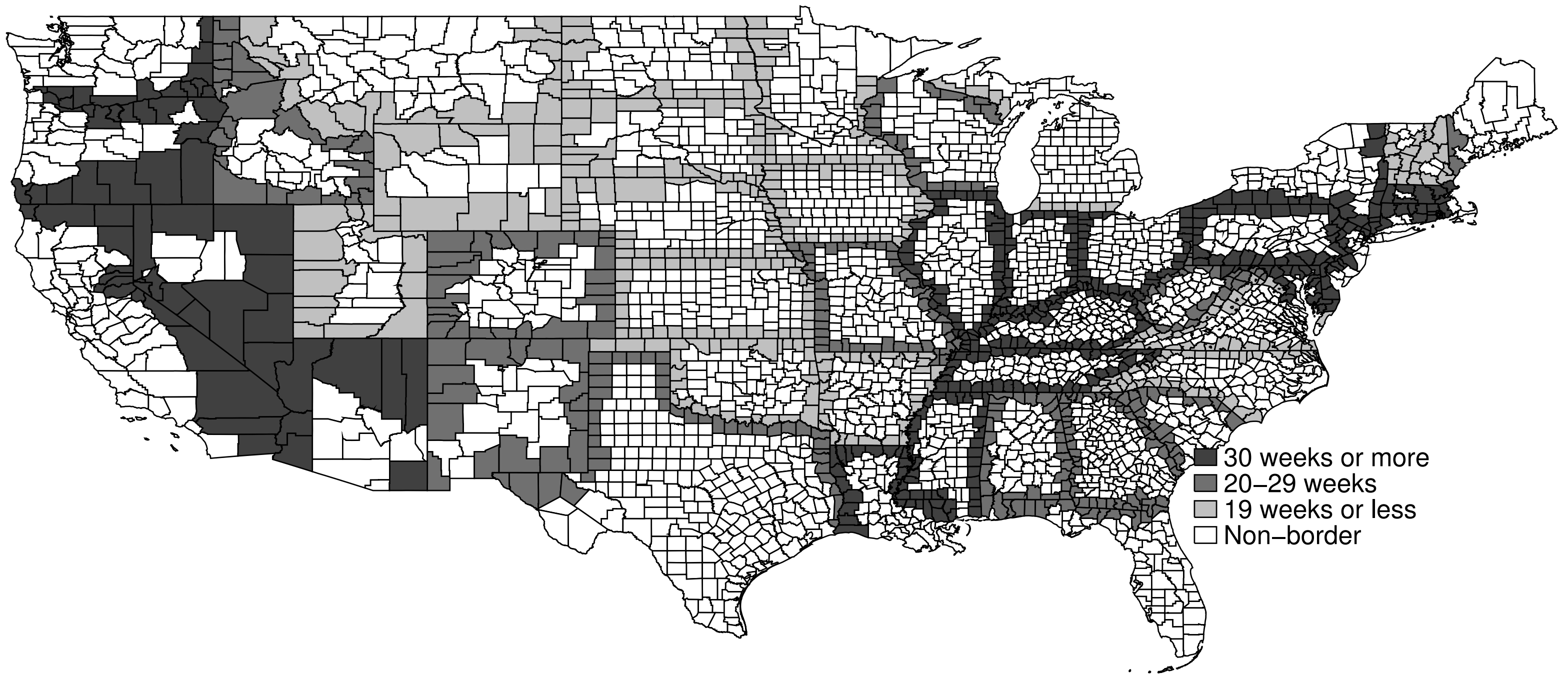
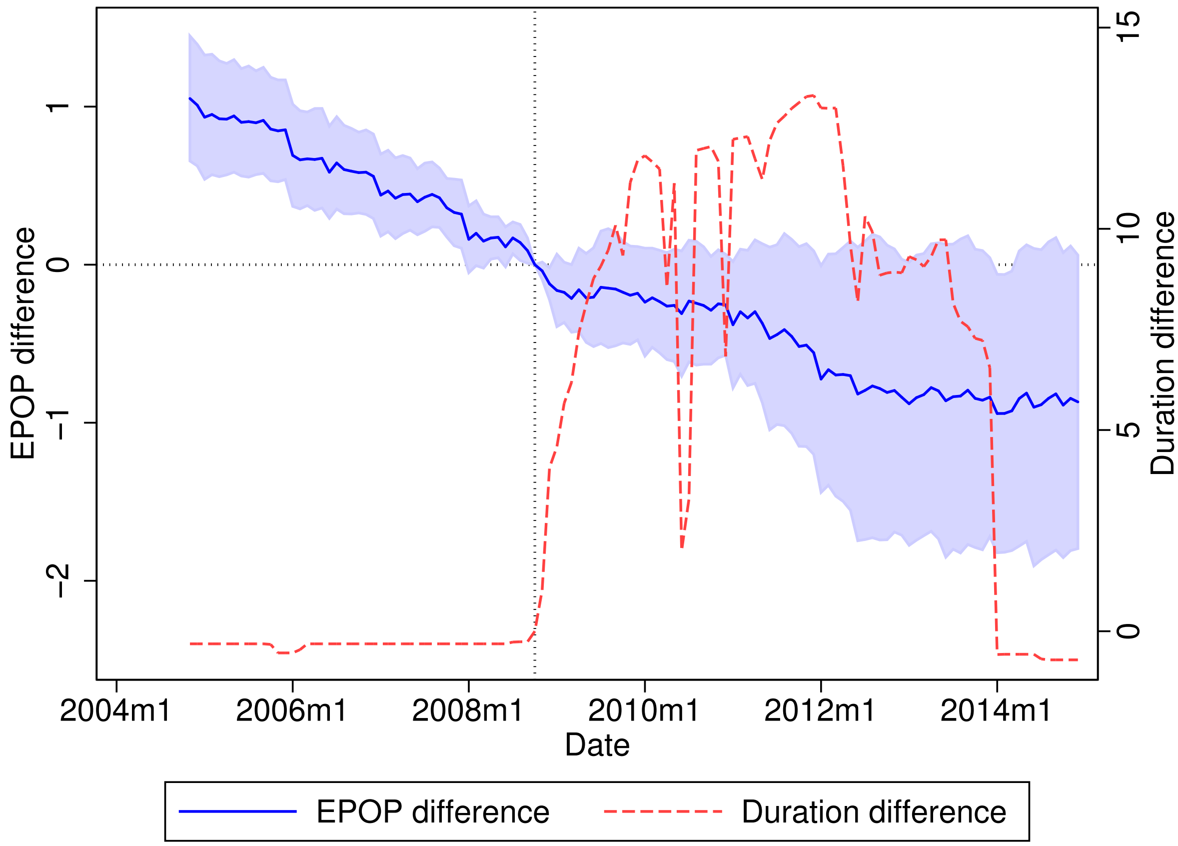
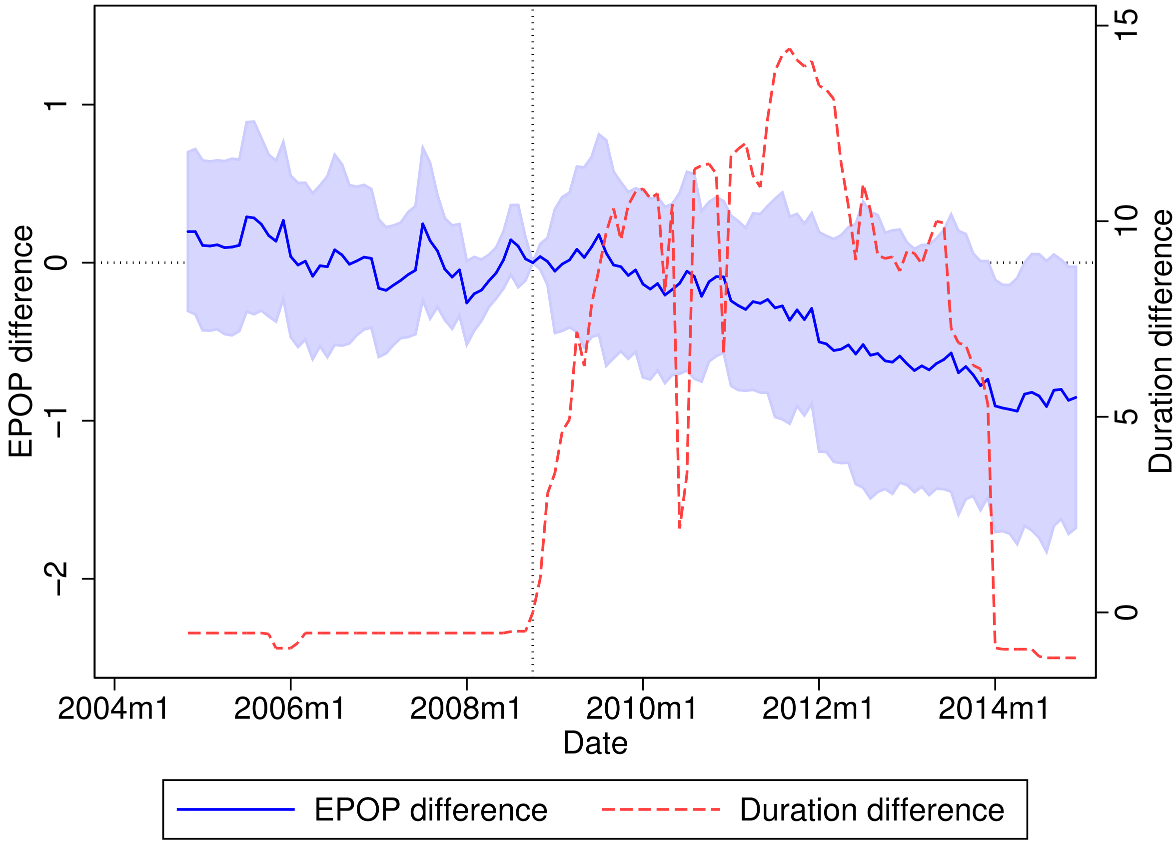


Figure 3: Evolution of EPOP and UI benefit duration differentials by average treatment intensity: without pair-period fixed effects



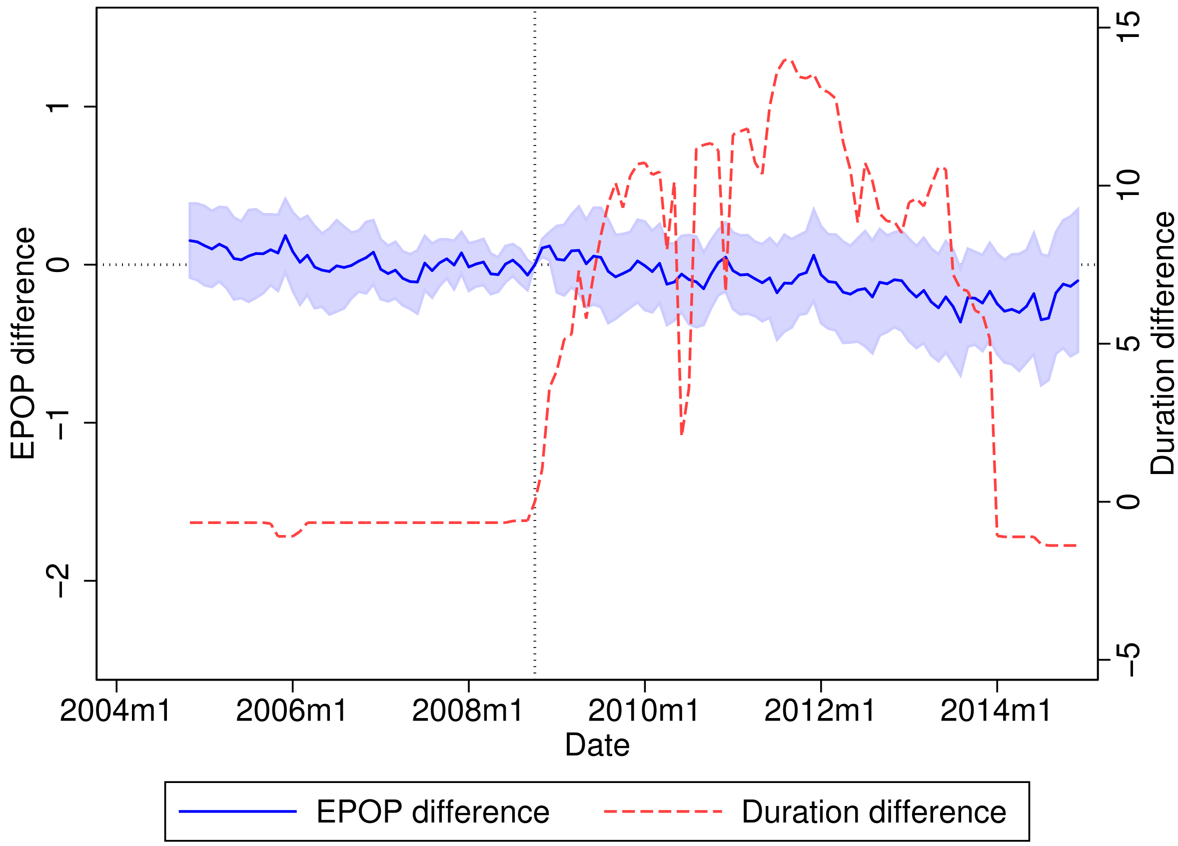
Notes: This figure plots (solid line, left axis) the set of β_s coefficients from the following regression: $E_{cpt} = \sum_{s=\tau_A}^{\tau_B} \beta_s treat_c \mathbb{1}\{t = s\} + \lambda_c + \nu_t + \epsilon_{cpt}$. E_{cpt} is the seasonally-adjusted ratio of total employment to population age 15+, 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 (2008m11-2013m12), minus average duration from the 12 months prior (2007m11-2008m10), divided by 10. The shaded region corresponds to the 95% 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 2008m10, 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.

Figure 4: Evolution of EPOP and UI benefit duration differentials by average treatment intensity: baseline border county pair sample



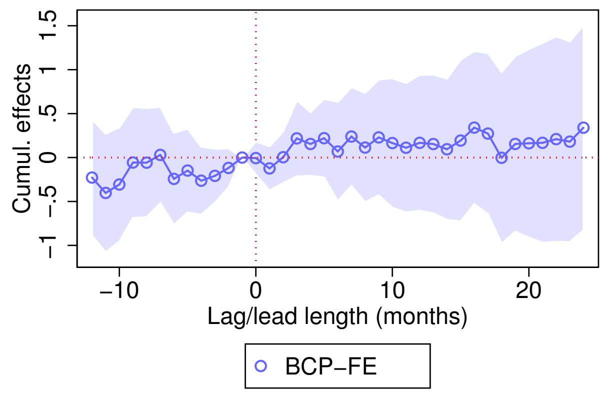
Notes: This figure plots (solid line, left axis) the set of β_s coefficients from the following regression: $E_{cpt} = \sum_{s=\tau_A}^{\tau_B} \beta_s treat_c \mathbb{1}\{t = s\} + \lambda_c + \nu_{pt} + \epsilon_{cpt}$. E_{cpt} is the seasonally-adjusted ratio of total employment to population age 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 (2008m11-2013m12), minus average duration from the 12 months prior (2007m11-2008m10), divided by 10. The shaded region corresponds to the 95% 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 2008m10, 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.

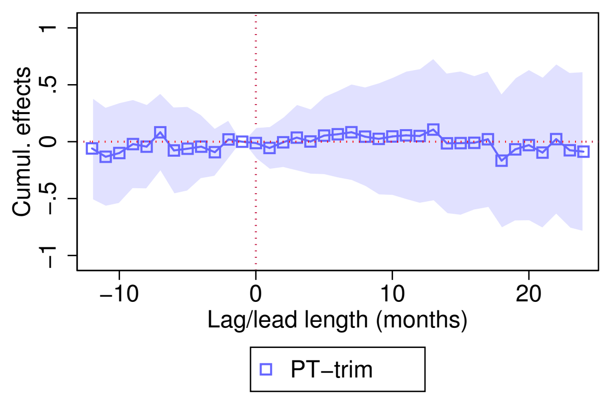
Figure 5: Evolution of EPOP and UI benefit duration differentials by average treatment intensity: PT-trimmed border county pair sample



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: $E_{cpt} = \sum_{s=\tau_A}^{\tau_B} \beta_s treat_c \mathbb{1}\{t = s\} + \lambda_c + \nu_{pt} + \epsilon_{cpt}$. 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 (2008m11-2013m12), minus average duration from the 12 months prior (2007m11-2008m10), divided by 10. The shaded region corresponds to the 95% 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 2008m10, 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.

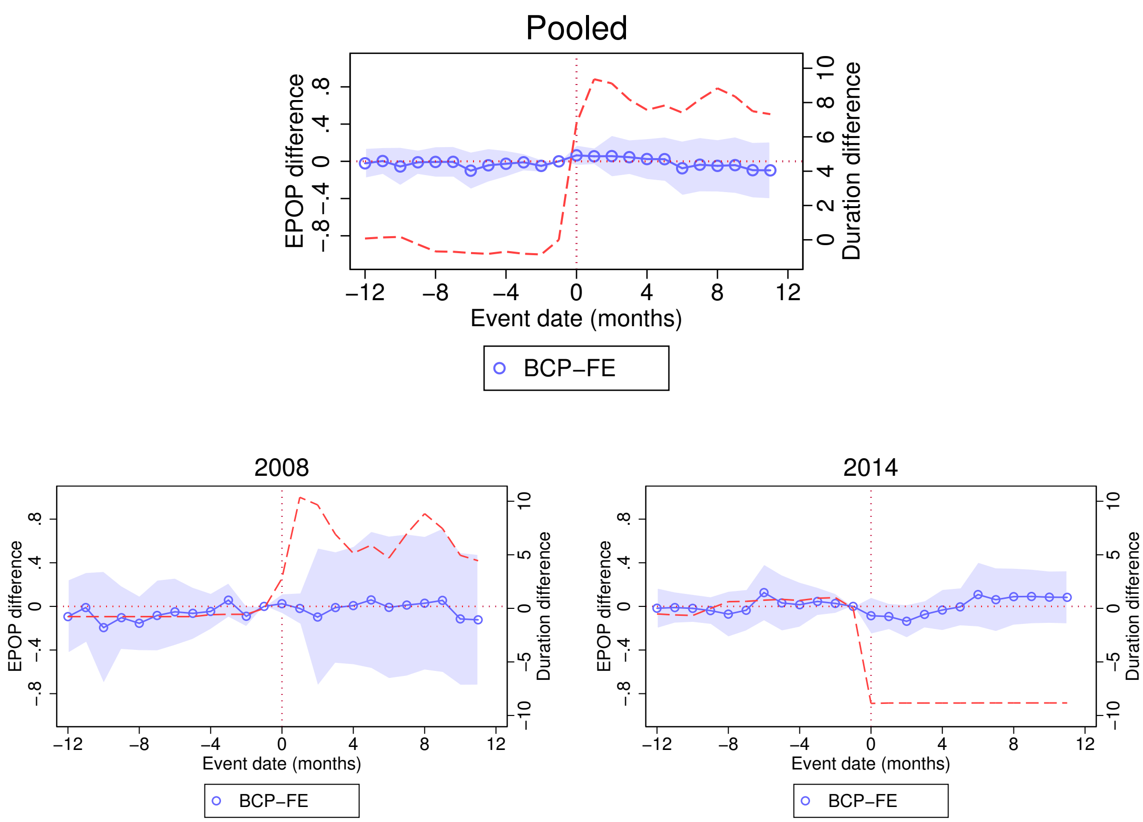
Figure 6: Cumulative response of EPOP from distributed lags specification: full sample regressions





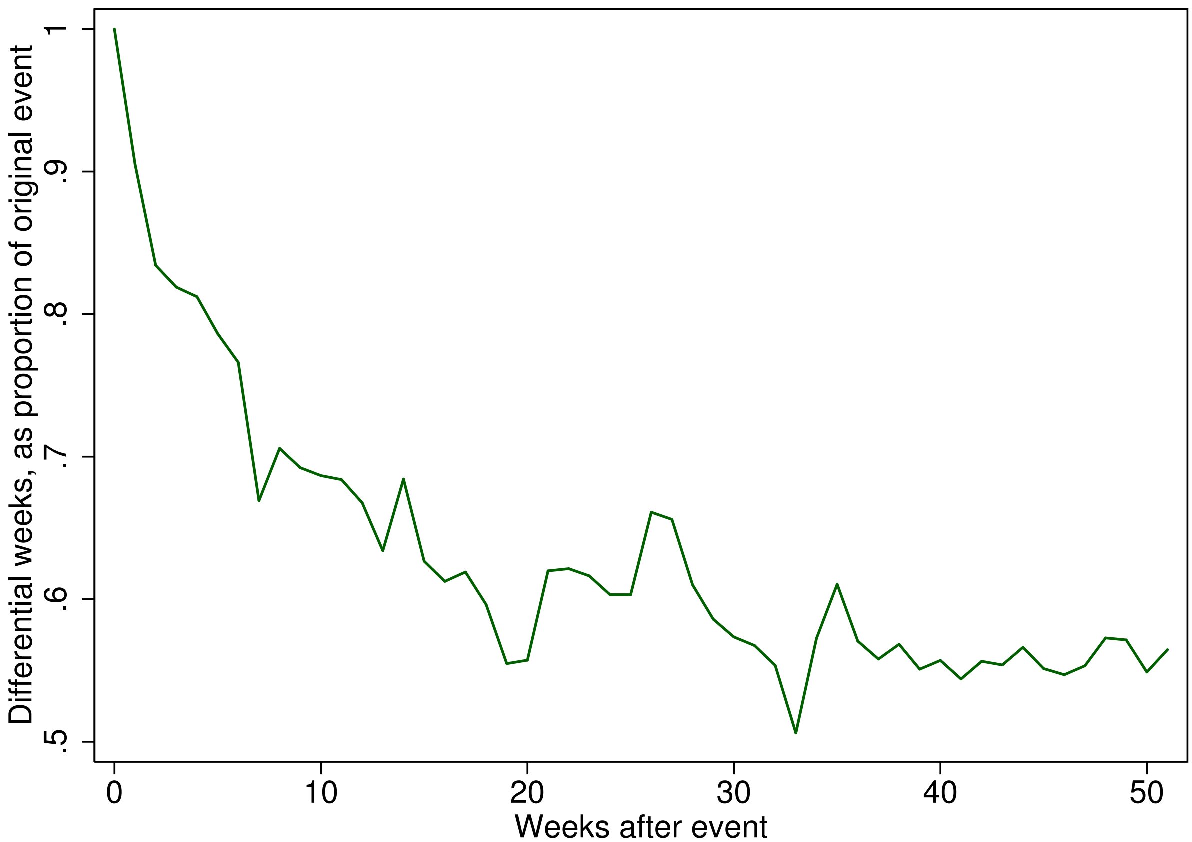
Notes: This figure reports the monthly cumulative response of EPOP from a 73 week increase in maximum UI benefit duration, centered around event date -1 whose cumulative response is defined as zero. The model is estimated on the full sample (2007m11-2014m12), using all border county pairs (BCPs) (hollow 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 2007m11-2014m12 (and thus duration data from 2005m11-2015m11). 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 1st through jth 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% confidence interval, robust to two-way clustering at the state and state-pair level.

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. The top panel uses pooled 2008 and 2014 samples, centered around event date -1 whose cumulative response is defined as zero. The bottom panel separately examines 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.

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.

Table 1: Summary statistics: High-treatment versus low-treatment counties in border county pair sample

	Baseline			PT-Trimmed		
	High	Low	p-val	High	Low	p-val
EPOP (A)	44.107 (17.044)	44.621 (15.110)	0.668	42.490 (14.545)	43.730 (13.422)	0.185
Private EPOP (A)	32.742 (14.468)	33.485 (14.074)	0.476	31.590 (12.653)	32.908 (12.500)	0.106
LAUS unemp. rate (A)	7.745 (2.358)	7.167 (2.403)	0.006	7.924 (2.182)	7.310 (2.233)	0.000
Population age 15+ (A)	81,415 (213,875)	71,441 (153,060)	0.211	96,552 (240,297)	80,856 (155,185)	0.015
Share white (B)	0.811 (0.182)	0.811 (0.177)	0.998	0.814 (0.179)	0.813 (0.176)	0.946
Share black (B)	0.085 (0.145)	0.086 (0.147)	0.966	0.090 (0.147)	0.088 (0.144)	0.834
Share hispanic (B)	0.067 (0.111)	0.059 (0.092)	0.491	0.060 (0.100)	0.053 (0.086)	0.468
Share H.S. grad (B)	0.569 (0.064)	0.567 (0.065)	0.724	0.567 (0.062)	0.565 (0.065)	0.691
Share college (B)	0.179 (0.078)	0.189 (0.086)	0.010	0.182 (0.081)	0.191 (0.087)	0.001
Median h.h. income (B)	42,645 (11,459)	43,535 (12,127)	0.198	42,898 (11,937)	43,969 (12,775)	0.119
New mortgage debt p.c. (A)	3.386 (3.092)	3.586 (2.877)	0.447	3.488 (2.874)	3.687 (2.924)	0.435
Share in cities 50k+ (C)	0.190 (0.331)	0.196 (0.331)	0.759	0.214 (0.345)	0.230 (0.348)	0.396
Min. weeks of UI elig.	24.470 (3.495)	24.631 (3.199)	0.718	24.436 (3.511)	24.787 (3.030)	0.494
Max. weeks of UI elig.	96.105 (6.674)	86.996 (13.320)	0.000	96.277 (6.560)	87.659 (12.622)	0.000
Pairs w/ different avg treatment	1131	1131		848	848	
Pairs w/ identical avg 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 2008m11-2013m12 period, minus average duration over the 2007m11-2008m10 and 2014m1-2014m12 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.

Table 2: Main Estimates: Effect of UI benefit duration on EPOP using full sample and event study specifications

	(1) BCP-FE	(2) PT-Trimmed
Full sample		
OLS Estimate	0.430	0.180
	(0.471)	(0.268)
County pairs	1161	870
Observations	199692	149640
Pooled sample (IV)		
IV estimate	0.143	0.253
	(0.974)	(0.650)
First stage coef.	0.847	0.842
	(0.052)	(0.052)
F stat.	[262.2]	[262.3]
County pairs	1161	870
Observations	108000	80928
2008 sample (IV)		
IV estimate	0.549	1.344
	(2.541)	(1.253)
First stage coef.	0.717	0.726
	(0.111)	(0.114)
F stat.	[41.3]	[40.3]
County pairs	1161	870
Observations	55728	41760
2014 sample (IV)		
IV estimate	-0.024	-0.214
	(0.568)	(0.523)
First stage coef.	0.915	0.903
	(0.046)	(0.044)
F stat.	[392.6]	[423.8]
County pairs	1089	816
Observations	52272	39168

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 the first panel use OLS estimated over the 2007m11-2014m12 period. Regressions in the remainder of the table are estimated on subsamples using instrumental variables. The instrument z_{ct} is defined as follows. From 2007m11-2008m10, z_{ct} is equal to the duration available immediately prior to the implementation of UCEA; from 2008m11-2009m10, z_{ct} is equal to the duration available immediately *after* the implementation of UCEA. From 2013m1-2013m12, z_{ct} is equal to the duration available immediately prior to the expiration of EUC; from 2014m1-2014m12, z_{ct} is equal to the duration available immediately after EUC expiration, before any changes in regular benefits took effect. Estimates in the second panel pool the 2007m11-2009m10 and 2013m1-2014m12 samples and replace county fixed effects with county-by-subsample fixed effects. Estimates in the third panel use data from 2007m11-2009m10; estimates in the fourth panel use data from 2013m1-2014m12. In the IV specifications, first stage coefficients and standard errors are also reported. Standard errors are reported in parentheses and first stage F-statistics in square brackets. Standard errors are clustered two-way at the state and state-pair level.

Table 3: Robustness of the effects of UI benefit duration on EPOP: choice of sample period

	(1) BCP-FE	(2) PT-Trimmed
2007m11-2014m12	0.430	0.180
	(0.471)	(0.268)
	N = 199692	N = 149640
2006m11-2014m12	0.142	0.107
	(0.455)	(0.272)
	N = 227556	N = 170520
2005m11-2014m12	-0.088	0.030
	(0.445)	(0.277)
	N = 255420	N = 191400
2004m11-2014m12	-0.330	-0.062
	(0.457)	(0.286)
	N = 283284	N = 212280
County pairs	1161	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 the top panel of Table 2. Standard errors are clustered two-way at the state and state-pair level.

Table 4: Additional robustness checks on the effects of UI benefit duration on EPOP

	Full sample OLS		**Pooled sample IV**	
	(1) BCP-FE	(2) PT-Trimmed	(3) BCP-FE	(4) PT-Trimmed
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(emp)	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 1st row correspond to the estimates in the top two panels of Table 2. The estimates in the 2nd row replace (total) EPOP with the ratio of private employment to population age 15+. In the 3rd row, we trim the set of border county pairs based on the level of correlation between county EPOP and state EPOP over the period 2004m11-2008m10 (see text for details). The 4th row controls for county-specific linear trends. The 5th row recodes the periods in 2010 when EUC lapsed by assigning EUC values during these lapses as equal to their prior value. The 6th row uses quarterly data instead of monthly (and estimates over the 2007q4-2014q4 period). The 7th row uses EPOP derived from the QWI (at the quarterly level) instead of the QCEW. The 8th row uses seasonally-unadjusted data. The 9th row includes counties without full EPOP data for each month, which we drop by default. The 10th and 11th row 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 $(\frac{99}{26}^{\hat{\beta}} - 1)\bar{E}$, where $\bar{E}$ is the mean EPOP level in the given sample. The 12th row uses a modified version of the instrument z_{ct} which 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 dropped in the 2014 subsample. The 13th row completely drops all NC BCPs. The 14th row keeps all North Carolina BCPs. The 15th row keeps NC BCPs but redefines the instrument for NC counties (see text for details). The 16th row drops county-pairs whose population centroids are greater than 100km apart. The 17th row uses the "hinterland" pairs rather than the border pairs (see text for details). Cells which are not applicable in the given sample, or which provide estimates that are mechanically equal to the baseline estimates, are left blank. Standard errors are clustered two-way at the state and state-pair level.

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); $\Delta E_{\text{MICRO}} = \beta_{\text{MICRO}} \times L$, 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; $\Delta E_{\text{MACRO}} = \beta_{\text{MACRO}} \times P$, where P is the population and β_{MACRO} is a direct estimate of the aggregate change in EPOP (from column 2 in the top panel of Table 2). The gap in ΔE is calculated as the difference between ΔE_{MACRO} and ΔE_{MICRO}.

Unemployment Insurance Generosity and Aggregate Employment

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

Online Appendix

A Online Appendix A: EB and EUC programs

Extended Benefits (EB)

Historically, when not in recession, most U.S. states have provided a maximum of 26 weeks of unemployment insurance to job-losers. At the onset of the Great Recession, in 2008, only two states offered more than 26 weeks of regular benefits. Massachusetts had a maximum of 30 weeks of UI benefits and Montana had a maximum of 28 weeks and no states offered less than 26 weeks.[51]

Since Congress created the Extended Benefits (EB) program in 1970, maximum benefit lengths increase automatically when unemployment is high and growing. At a minimum, in states where the Insured Unemployment Rate (IUR) exceeds 5%, and the IUR is at least 1.2 times the IUR in the previous two years, claimants are eligible for 13 additional weeks of UI after the expiration of regular benefits.[52] The same law also provides two optional "triggers," which can be adopted by states at their own discretion. The first trigger provides for 13 weeks of EB for states whose IUR exceeds 6% (regardless of the change in the IUR over time). The other optional trigger is based on the Total Unemployment Rate (TUR): the trigger provides for 13 weeks of EB when both (1) the TUR exceeds 6.5% and (2) the current TUR is at least 1.1 times its value in the prior two years. States adopting this second trigger must provide 20 weeks of EB when (1) the TUR exceeds 8%, subject to the same growth-over-time requirement.[53] States can adopt zero, one, or both optional triggers, but no more than one trigger can be "on" at any point in time, meaning that the number of weeks of EB is capped at 20.

Normally, the costs of EB are shared equally between the federal and state governments. As a result, many states did not have statutes activating the optional EB triggers at the onset of the Great Recession.

[51] Not all claimants are eligible for the maximum number of weeks of benefits. In most states, individuals with relatively weak recent labor force attachment are eligible only for a fraction of the maximum weeks of benefits. Throughout this paper, we abstract from this complication by focusing on the maximum UI duration. Our estimates, therefore, can be seen as an intention to treat effect. Johnston and Mas (2018), using micro-data from Missouri, find that approximately 70% of UI claimants had sufficient labor force attachment to be eligible for the full 26 weeks of regular benefits from 2003-2013.

[52] The Insured Unemployment Rate (IUR) is, roughly, the ratio of current regular UI claimants to the number of UI-covered jobs. The Total Unemployment Rate (TUR) is the usual "unemployment rate": i.e., the ratio of unemployed persons to persons in the labor force.

[53] From December 2010 through the end of 2013 (a period in which the unemployment rate remained high but was generally not growing), states were allowed to apply a three-year lookback period instead of a two-year lookback period for the purpose of determining growth over time.

After the passage of the American Recovery and Reinvestment Act (ARRA), the federal government paid for the full amount of EB extensions. Some states (mostly deeply conservative ones) nonetheless declined to activate the optional triggers. For example, while Mississippi had a TUR of well over 8% continuously from January 2009 through October 2016, peaking at over 11% in 2010, they were never eligible for EB because the IUR never went above 5.6% and the state declined to enact the optional triggers. Thus, different states had different numbers of weeks of EB in part due to differences in the state unemployment rates and in part due to state policy differences. The federal government maintained its full support of EB until the end of 2013 when it returned to the default equal cost sharing rule.

Emergency Unemployment Compensation (EUC)

In response to the first signs of a weakening labor market, on June 30, 2008, Congress and President Bush created the Emergency Unemployment Compensation (EUC) program. At first, EUC provided for 13 additional weeks of benefits for all UI-eligible unemployed workers.[54] The Unemployment Compensation Extension Act of 2008 was then signed into law by President Bush on November 21, 2008. It augmented the EUC program while also creating the first differences across states in their access to the EUC extensions. It authorized 20 weeks of EUC for all states (an increase from 13) and an additional 13 weeks for those with a total unemployment rate exceeding 6%.[55] These additional weeks were organized into "tiers": Tier 1 corresponded to the first 20 weeks of EUC, while Tier 2 corresponded to the baseline 20 weeks plus an additional 13 weeks. During this period, a state with 26 weeks of regular benefits could qualify for up to 79 weeks total of benefits. Then, on November 6, 2009, the Worker, Homeowner, and Business Act of 2009 further increased maximum UI duration. Tier 1 remained in place. However, Tier 2 was increased from 13 to 14 weeks and extended to all 50 states. The law also added Tier 3, providing 13 additional weeks to states with a TUR of greater than 6%, and Tier 4, providing 6 additional weeks for states with a TUR of greater than 8.5%. After the passage of this law, states had access to a maximum of 99 weeks of benefits. This schedule remained in place, with the exception of temporary lapses, until early 2012, when Congress enacted laws that slowly began to phase out EUC.[56]

[54]To be more precise, this legislation—and all subsequent legislation related to EUC—provided for increases in benefit lengths equal to the lesser of (1) a specified number of weeks or (2) a fraction of the number of weeks of regular benefits. For the initial legislation in June 2008, the specified number of weeks was 13 and the fraction of the number of weeks of regular benefits was 50%. For the vast majority of states that had regular benefits greater than or equal to 26, the specified number of weeks was the binding factor. For those states with fewer than 26 weeks of regular benefits, the percentage of regular benefits was always binding. In this paper, we code the weeks available under EUC exactly as specified in the law; however, in the discussion that follows, we discuss only the specified number of weeks, which applies to states with at least 26 weeks of regular benefits.

[55]A state could also have become eligible for 33 weeks with a sufficiently high IUR; in practice, the IUR trigger was never binding.

[56]There were four lapses in EUC that occurred in 2010, arising due to political disagreements regarding the extension of the program. The longest such lapse lasted from May 30, 2010 to July 18, 2010. In each of the lapses, beneficiaries were paid

On February 22, 2012, Congress passed and the President signed The Middle Class Tax Relief and Job Creation Act of 2012 which slightly lowered the generosity of the EUC in a gradual way, first starting on May 27, 2012, and then again on September 2, 2012. By September 2, 2012, Tier 1 had been scaled back to 14 weeks and was still available to all states. Tier 2 remained at 14 weeks but again became available only to states with a TUR of greater than 6%. Tier 3 was scaled back from 13 to 9 weeks and the state TUR threshold was raised to 7%. Finally, Tier 4 was increased to provide 10 extra weeks for states with a TUR of above 9%. The program finally came to an end at the end of December 2013.[57] In total, over the Great Recession, individuals in qualifying states received up to 99 weeks of unemployment insurance. Compared to the baseline of 26 weeks, this is an increase of 73 weeks; so the maximum UI benefit duration in some qualifying states increased by almost 300%.

retroactively for any weeks of missed payments. Furthermore, during these lapses, the funding rules for EB reverted to their pre-ARRA levels, which led many states to suspend EB payments during these lapses as well.

[57]Upon the expiration of EUC at the end of 2013, EUC beneficiaries immediately stopped receiving benefit payments. Prior to the final expiration, however, the phase-out was more gradual. If a state "triggered-off" a certain tier, people who had already qualified for a given tier were allowed to finish that tier but were not allowed to move to the next tier. One exception, discussed in the main text, is North Carolina, which lost access to all EUC money as of July 1, 2013. In our econometric specifications, our duration variable is the maximum duration available in a given month for a new entrant into unemployment. Thus, we do not distinguish between gradual phase-outs and sudden benefit cessations.

B Online Appendix B: Comparison with HKMM and HMM

The results in this paper are quite different than the results in Marcus Hagedorn, Fatih Karahan, Iourii Manovskii and Kurt Mitman (2015) (which studies the effect of UI from 2005 to 2012) and the results in Marcus Hagedorn, Iourii Manovskii and Kurt Mitman (2016) (which studies the effect of EUC expiration at the end of 2013). Similar to this paper, both HKMM and HMM use border county pairs for their estimation. There are differences in data, in econometric specification, and in sample definitions between our paper and these two studies. Some differences are minor, while others are quite important. In this online Appendix, we expand upon the discussion in the main text to offer additional details regarding differences between our methods and results and those of HKMM and HMM.

Comparison to HKMM

In this section, we compare our full sample estimates from the baseline BCP sample to the baseline estimates of HKMM. The HKMM estimation equation is as follows, where data for a given pair p at time t has already been spatially differenced (after taking logs):

$$\ln(u_{pt}) - \beta(1 - s_t)\ln(u_{pt+1}) = \alpha * \ln(D_{pt}) + \lambda_p' F_t + \epsilon_{pt} \tag{B1}$$

Here, u_{pt} is the unemployment rate from LAUS,[58] β is the discount factor (equal to 0.99), s_t is the separation rate, D_{pt} is the same measure of maximum benefit lengths that we use, and $\lambda_p' F_t$ are interactive effects. Thus, the dependent variable is a quasi-forward difference (QFD) of the log of the unemployment rate. They then calculate the total effect of UI on unemployment by considering the steady state ($u_{pt} = u_{pt+1}$) impact of a persistent increase in D_{pt}. In the steady state, $\ln(u_p) = \frac{\alpha}{1-\beta(1-s)}\ln(D_p)$. Therefore, HKMM's headline claim comes from multiplying their main estimate by a factor $\frac{1}{1-\beta(1-s)}$, which is approximately equal to 10. They perform their estimation over the period 2005q1-2012q4.

Our full sample BCP-FE estimation strategy is different from HKMM in six distinct ways. These differences are: (1.) we do not transform our dependent variable using quasi-forward-differencing, (2.) we use employment data from the QCEW rather than unemployment data from LAUS, (3.) we estimate the results using monthly data from 2007m11-2014m12, instead of quarterly data from 2005q1-2012q4, (4.) we control for differences across county pairs using a fixed effects model rather than the Bai (2009) interactive fixed

[58]The LAUS data used by HKMM has been substantially revised since they accessed it. We have estimated the models using both the pre-revision version of the LAUS data used by HKMM and the more recent, revised version of the data. We have found both versions of the data give similar results in the HKMM specifications. We use pre-revision data throughout the discussion of HKMM.

effects model, (5.) we use levels instead of logs, and (6.) we restrict ourselves to a balanced panel, throwing out 10 small counties which did not report county-level employment at least once during our sample period due to disclosure issues.

Appendix Table B1 describes the impact of each of these six steps. Because different specifications have different dependent variables, and because the implied effect is not equal to the coefficient in some specifications, we standardize each specification into an implied effect of the 26-to-99 week expansion on EPOP.[59] We "translate" between implied effects on the unemployment rate and implied effects on EPOP by using the total peak-to-trough impact of the Great Recession. We measure this peak-to-trough impact using the unweighted average of counties in our border-pair sample. In particular, in this sample, EPOP fell from 44.3% to 41.2% and the unemployment rate increased from 4.8% to 9.7%. So, if one estimation suggests that the impact of the 26-to-99 week expansion was 3 percentage points of unemployment, we would convert that specification's estimate into an EPOP effect of $3 \times (\frac{41.2-44.3}{9.7-4.8}) \approx -1.9$ percentage points.

Appendix Table B1 analyzes one-off changes either starting from the HKMM specification (column 1), or moving to our specification (column 2). The first row begins with reporting the estimates: our replication of the HKMM estimates, joint with Dieterle, Bartalotti and Brummet (2020) and discussed in Online Appendix D, suggest that the UI benefit expansion from 26 to 99 weeks has an implied EPOP effect of -2.66, which more than 85% of the decrease in EPOP during the Great Recession within our sample. This corresponds to a coefficient estimate of 0.051, while HKMM report a very similar estimate of 0.049. We find that this estimate is statistically significant, as HKMM do. In contrast, the point estimates for the full sample BCP-FE estimates in this paper suggest that the decline in EPOP would have been about 10% *greater* without the UI expansions, though this is not distinguishable from zero.

The next five rows report the marginal impact of each of the five steps. In column 1, we show what happens when the step reported in the row is added starting with the HKMM specification. In column 2, we show what happens when this step is added to our specification. Finally, in column 3, we consider *all* possible transition paths between HKMM's estimates and our estimates, and report the average marginal contribution of each of the steps, across all of these transition paths.[60]

The key findings are as follows. Quasi-forward differencing, the use of the LAUS unemployment data as opposed to the QCEW employment data, and sample alignment are all consequential choices. In contrast,

[59] Importantly, we scale up the estimates in QFD specifications by $1/[1-\beta(1-s)]$, as HKMM do.

[60] We do not consider the step of switching from logs to levels in column 1, because the quasi-forward-differencing is motivated by theory which requires the data to be in logs. With quasi-forward-differenced data in levels, it is neither clear what we are measuring, nor what the total effect of UI on employment would be. For the same reason, we do not consider adding quasi-forward-differencing to our specification in column 2 (which is in levels). In addition, when calculating the averages in column 3, we discard transition paths that involve using quasi-forward-differenced data in levels. In the end, we estimate 48 models with all allowable combinations of the five sources of differences; we then take 360 paths (equal to 6! paths with 1/2 thrown out because eliminating quasi-forward differencing happens after the logs to levels conversion) between the HKMM and BDGK estimates, and calculate the contribution of each of these six factors averaged across these 360 paths.

the use of interactive fixed effects as opposed to linear fixed effects, the use of logs versus levels, and the use of a balanced as opposed to unbalanced panel of counties are not consequential choices.

Column 1 shows that, starting from the HKMM estimate, switching from the LAUS unemployment rate, or getting rid of quasi-forward differencing, dramatically reduces the HKMM estimates in magnitude towards zero. In particular, just switching from the LAUS unemployment rate to the QCEW EPOP (as shown in Row 4) changes the estimates to $-2.661 + 1.120 = -1.541$, suggesting the UI benefit expansion explained around 50% of the fall in EPOP rather than 85% as implied by HKMM's estimates. Similarly, removing quasi-forward differencing (Row 2) changes the estimates to $-2.661 + 2.618 = -0.043$ percentage points of EPOP. Column 2 shows that use of the LAUS unemployment rate also leads to a (mistaken) suggestion of job loss when we start from our specification, although the impact of this is more modest. Starting from our BCP-FE specification, when we use the LAUS unemployment rate as the outcome, the translated result suggests the UI benefit expansion led to a change in EPOP equal to $0.430 - 1.133 = -0.703$, just under a quarter of the overall change during the Great Recession. When we average the incremental contribution of these two steps across all permissible paths going between the HKMM specification and ours (in column 3), we find that dropping quasi-forward differencing increases the estimates by around 1.30 percentage points of EPOP (about 40% of the change in unemployment rate during the Great Recession), while switching the outcome from LAUS unemployment rate to QCEW based EPOP increases the estimate by about 0.67 percentage points of EPOP.

Aligning our sample period also has a meaningful impact. The HKMM sample of 2005q1-2012q4 starts and ends earlier than our sample of 2007m11-2014m12. Averaged across all sample paths, moving from HKMM's sample to ours adds 0.91 percentage points of EPOP to the estimate. As we showed in **Table 3**, while the baseline BCP-FE approach greatly reduces the pre-existing trend, it does not completely remove it. Use of an earlier start date, as well as an end date prior to the phase-out of differential UI benefits across state borders, can produce a more negative estimate in the presence of such trends. We find that use of this altered sample period leads to somewhat smaller magnitudes of estimates, reducing the impact of the policy by around 0.817, 1.572, and 0.915 percentage points of EPOP in columns 1, 2, and 3, respectively.

In contrast, the use of Bai (2009) interactive effects versus fixed effects, the use of logs versus levels, and the use of a balanced panel make fairly small contributions in explaining the difference between our two sets of estimates.

This analysis shows that (1) changing the sample period (and frequency) from HKMM's specification to ours, (2) eliminating quasi-forward-differencing, and (3) changing the dependent variable from the LAUS unemployment rate to QCEW EPOP all reduce the implied negative impact of UI on employment, by 0.67 to 1.30 percentage points of EPOP when averaged over all possible paths. We next discuss our justification for making the specification choices that we do.

Quasi-Forward Differencing

HKMM derive **Equation (B1)** by considering a search-and-matching framework where the rate of vacancy posting or firm job creation depends on a firm's expectation about future wages. Since unemployment insurance puts upward pressure on wages, an increase in benefits would reduce the expected profits of the firm and lead to a reduction in job creation. Because expectations about *future* benefit changes can affect employment *today*, HKMM make the point that an empirical approach that only relates current employment to current or past policy changes would be misspecified. In order to capture these anticipation effects, HKMM use a quasi-forward-differencing procedure. Their argument is as follows: the value of an employee to an employer is equal to the current-period flow profits, plus $\beta(1-s)$ times the expected value of the employee tomorrow (since the value of a vacant job is driven to zero by free entry). Therefore, HKMM argue, we can isolate the impact of UI on current-period flow profits by considering the quasi-forward difference of the unemployment rate (which they consider to be proportional to current period flow profits, in logs). The theory predicts that, in the case of an increase in generosity that was a surprise and immediately known to be persistent, firms would move from a low-unemployment steady state to a high-unemployment steady state, according to the equation $\Delta \ln(u_p) = \frac{\alpha}{1-\beta(1-s)} \Delta \ln(D_p)$.[61] As we noted above, this choice is quite important—removing forward differencing essentially erases the entirety of their effect even in their sample.

We are generally less favorable toward the use of quasi-forward differencing for several reasons. This model-driven approach relies on strong parametric assumptions—most notably that labor demand is well-characterized by the vacancy-posting problem captured in the model. Unfortunately this results in an empirical approach that is very sensitive to misspecification. For example, if an increase in UI generosity (D_{pt}) tends to be associated with a decrease in future unemployment (u_{pt+1}) in the data, then the estimated coefficient α will be positive. Such a pattern could also be consistent with a Keynesian aggregate demand effect that operates with a small delay. That is, if an increase in benefits in one period leads to increased aggregate demand and lower unemployment in the next period, the HKMM strategy would find that UI *increased* the unemployment rate, when in fact the opposite occurred. This problem is illustrated in **Appendix Figure B1**. Second, as a practical matter, the size of the final estimate is sensitive to assumptions in the model required for translating a flow result to a steady state effect, and in the exact magnitudes of separation and discount rates. Both the heavy dependence on a specific model and the inability to distin-

[61] Here α is the regression coefficient, β is the discount factor, s is the probability that the job ends, u is the unemployment rate, and D is the number of weeks of UI benefits.

guish between alternative explanations make quasi-forward differencing an unattractive strategy from our perspective.

Instead, our preferred strategy is to capture the dynamics in a less model-driven and a more transparent manner using distributed lags. That specification directly estimates employment changes around benefit duration innovations, allowing us to assess possible pre-existing trends, anticipatory effects, and delayed or slow moving response within the window. As we discussed in Section IV.A, we find no evidence of significant anticipation effects in the 12 months prior to benefit changes. The lack of any anticipation effect raises questions about the value of quasi-forward differencing the outcome, especially given the drawbacks discussed above.

LAUS versus QCEW

HKMM predominantly use the LAUS employment data rather than the QCEW employment data to compute county-level measures of employment.[62] Importantly, the LAUS data is partly model-based. In particular, while the LAUS data uses actual movement to unemployment based upon UI claims, they do not observe those entering (or re-entering) the labor force. Therefore, the county-level estimates for unemployment are based on state-level data on labor force entry and re-entry—something BLS states explicitly in their online manual (http://www.bls.gov/lau/laumthd.htm):

> "The second category, "new entrants and reentrants into the labor force," cannot be estimated directly from UI statistics, because unemployment for these persons is not immediately preceded by the period of employment required to receive UI benefits. In addition, there is no uniform source of new entrants and reentrants data for States available at the LMA [labor market area] level; the only existing source available is from the CPS at the State level. Separate estimates for new entrants and for reentrants are derived from econometric models based on current and historical state entrants data from the CPS. These model estimates are then allocated to all Labor Market Areas (LMAs) based on the age population distribution of each LMA. For new entrants, the area's proportion of 16-19 years population group to the State total of 16-19 years old population is used, and for reentrants, the handbook area's proportion of 20 years and older

[62] They do report results using the log employment from the QCEW and QWI as a robustness check, in columns 3 and 4 of Table 5. The log employment result, -0.03, would imply that the 26-99 week expansion of UI caused a reduction of employment by 3.9%, which would translate to about 1.6 percentage points of EPOP. This is about 40% less than implied EPOP effect of HKMM's main result, consistent with the average marginal effects reported in **Appendix Table B1**. The log employment results from the QWI are modestly larger.

population to the State total of 20 years and older population is used."

The use of state-level information in estimating county-level unemployment rates is problematic for a border discontinuity design. The border county design attempts to purge reverse causation present at the state level by using more local comparisons. Use of state-level information raises the possibility of finding a (spurious) discontinuity in the measured unemployment rate across the state borders even when there is no such discontinuity in reality.

The QCEW data are based on administrative payroll records provided to the BLS by states, which protects against finding spurious discontinuities. Moreover, the QCEW data includes around 98% of all formal sector workers, making them very close to the true total employment counts in these counties. For these reasons, we consider the QCEW to be the preferred data source for county-level employment. When the results using the QCEW and LAUS data differ non-trivially—which they do in this case—the QCEW findings are much more likely to be accurate.

Sample Alignment

HKMM's sample goes from 2005 through 2012 and uses quarterly data. By contrast, our main specification uses monthly data, starts in 2007m11, and goes through 2014m12. Using quarterly versus monthly data has virtually no impact. For our preferred specification, for example, changing to quarterly data increases the standard errors by a little more than 0.04 and increases the the mean estimate by 0.02 (see **Table 4**). Though that represents a 7% increase, since the baseline estimates are small to start with, the impact is quite small. Switching the time period of estimation from 2005-2012 to 2007m11-2014m12 does make a difference. First of all, as we discussed in **Section IV.C**, the 2007m11-2014m12 sample exhibits a fairly symmetric rise and then fall in treatment intensity, orthogonalizing possible trends. Moving to the 2005-2012 sample makes this less so. As can be seen in **Figure 4**, the 2005-2012 period is largely a period of (1) increasing benefit duration and and (2) decreasing relative employment on the high-treatment side of the border. After 2012, the high-treatment side of the border starts to experience a relative decline in duration, while continuing its relative decline in employment. This is in part due to federal policy changes and in part due to differential changes in unemployment levels. Thus, it is not surprising that adding 2013 and 2014, and removing 2005 to 2007m10, has a noticeable positive impact on the UI duration impact upon employment.

Furthermore, we note that the choice of sample date matters little for the PT-trimmed sample. **Table 3** shows that the estimates in the full sample BCP-FE specification fall from 0.430 to -0.330 when the sample is changed from 2007m11-2014m12 to 2004m11-2014m12, while the estimates in the PT-trimmed sample fall only from 0.18 to -0.06. The IV estimates show a similar pattern, although the range is larger in both samples. This leads us to be confident that the large negative effects seen in full sample specifications with earlier start dates (and/or end dates) reflect endogeneity from pre-existing trends. Furthermore, since the 2007m11-2014m12 sample window effectively orthogonalizes these trends with treatment, we believe that our sample window provides for more reliable estimates than other sample windows, including HKMM's 2005q1-2012q4.

HMM comparison

HMM find that the expiration of EUC at the end of 2013 increased employment, though the implied effect of UI generosity is smaller than that of HKMM. Whereas the latter suggests that approximately 80% of the increase in unemployment during the Great Recession can be explained by the increase in benefit generosity, applying the coefficient estimates of HMM to the 26-to-99 week expansion would imply that UI policy can explain about one third. Scaled another way, HMM finds that the employment effect of the *expiration* is on the same order as total employment gains during 2014. HMM estimate a variety of different empirical models, all of which are motivated by a desire to exploit variation in UI benefits solely coming from the EUC expiration, while at the same time incorporating information over a longer period to formulate a counterfactual for the county-level employment which would have occurred had EUC not expired. Broadly, these specifications can be broken into two groups, which we call the "interaction term" models and the "event study" models.[63] We discuss each of them in turn.

The following is equivalent to HMM's "benchmark" interaction term model, where e_{ct} is log employment, measured either in the QCEW or LAUS:[64]

$$e_{ct} = \kappa[\ln(D_{ct})\mathbb{1}(t \leq 2013q3)] + \alpha[\ln(D_{ct})\mathbb{1}(t \geq 2013q4)] + \mu_c + \nu_{pt} + \gamma_c t + u_{cpt} \tag{B2}$$

That is, the model includes pair-period fixed effects, county fixed effects, as well as a county-specific time trend. The coefficient of interest is α, which measures the effect of duration on employment solely using variation from 2013q4 onward (i.e., from no earlier than the quarter immediately prior to expiration). The other independent variable, the log of benefit duration in periods prior to 2013q4, soaks up the effect of duration up to 2013q3; this ensures that, after taking out county fixed effects and county-specific linear trends, the model is comparing employment differences in 2013q4 to employment differences in all quarters

[63] The former correspond to models discussed in Sections 3 through 5 of HMM and the latter correspond to models discussed in Section 6 of HMM.

[64] We understand that HMM takes the spatial difference across pairs manually; as discussed above, this is equivalent to including a full set of pair-period fixed effects.

in 2014.

The first column of the top panel of **Table B3** shows HMM's estimate of this specification over the 2010q1-2014q4 period, as well as our replication. They estimate a coefficient of -0.0190, with a p-value of zero (to three decimal places) from a block bootstrap procedure. To place this in the context of our other estimates, this would translate into a -1.05 percentage point reduction in EPOP from a 26-to-99 week expansion of duration. While this is smaller than the corresponding estimate in HKMM, it is still substantial, representing about one third of the EPOP drop of the Great Recession; it would also imply that the expiration of EUC was responsible for increasing employment in 2014 by about 2 million jobs. When we estimate this equation using the LAUS data that they use on the county pairs in our sample, we estimate a very similar coefficient of -0.0200, with an analytical standard error (clustered at the state-pair level) of 0.0082,[65] which implies a p-value of about 0.015.[66] Since HMM accessed their data, the entire LAUS series has been redesigned by the BLS, largely to incorporate information from the American Community Survey rather than the Decennial Census.[67] The second column of the first panel shows our estimate from the same specification but with employment derived from the revised data. The coefficient falls in magnitude by three quarters to -0.0048 and becomes statistically indistinguishable from zero. Thus, when using the most recent version of the LAUS employment series, this specification no longer finds that the 2014 EUC expiration caused an employment boom.

HMM also estimate this model using log employment derived from the QCEW and find a modestly negative estimate of -0.0100. In our scale, this would translate to an EPOP effect of -0.558 percentage points from a 26-to-99 week expansion. When we estimate their model we obtain a similar coefficient of -0.0078, corresponding to an EPOP effect of -0.435.[68] While -0.558 is more negative than our 2014 IV specification (-0.024 in the full BCP-FE sample, or -0.214 in the PT-trimmed sample), the difference is at the bottom end of the range of estimates that can be generated using QCEW data from robustness checks on our main specifications. In results available upon request, we re-estimate our baseline 2014 BCP-FE IV specification using all combinations of the following specification choices: (1) using EPOP, log EPOP, or log employment as the dependent variable,[69] (2) using duration in logs or in levels as the independent variable of interest,

[65]In our baseline specifications, we cluster two-way at the state and state-pair level in order to account for any common state-level shocks (including mechanical correlation of errors for those counties that border multiple states). For the sake of this reconciliation exercise, we cluster at the state-pair level. Clustering at the two-way level in this specification increases the standard error to 0.0097.

[66]Our baseline sample includes 1,161 county pairs, and we drop an additional two pairs due to missing data in this specification. While our baseline specification studying the 2014 EUC expiration drops pairs in which either county is in North Carolina, we do not drop such pairs in this reconciliation exercise. HMM report using 1,175 pairs with full data. Such a discrepancy could arise due to reasonable differences in interpretation regarding, e.g., whether counties that touch only on a corner should be included as a "county pair."

[67]See http://www.bls.gov/lau/lauschanges2015.htm for details. We downloaded the current LAUS data on November 10, 2016.

[68]In our baseline specifications in this paper, we seasonally adjust the QCEW data as described in the text. For the sake of this reconciliation exercise, we use not-seasonally-adjusted data.

[69]We do not estimate a specification using employment in levels.

(3) keeping county pairs involving North Carolina or dropping them, (4) defining the instrument based on changes in duration immediately upon the EUC expiration, or defining it based on the change between average duration in 2013q4 and the average duration in 2014, (5) starting the sample in 2013q1 or 2013q4, and (6) using seasonally-adjusted or not-seasonally-adjusted data. After translating each estimate to its implied effect on EPOP in levels, we find that these 96 estimates range between -0.637 and 0.541. The EPOP-equivalent estimate from HMM specification using QCEW data (either -0.558 using their estimate or -0.435 using our replication) is within that range, though at the negative end. Furthermore, as with the LAUS specification, we find a lower level of statistical precision than HMM: our standard error of 0.0068 would mean that HMM's point estimate of -0.0100 would not be statistically distinguishable from zero at conventional levels.

HMM repeat their analysis with two variants of their benchmark model. First, they replace the county fixed effects and linear trends with interactive effects (Bai 2009) and estimate the model over the 2005q1-2014q4 period. Second, they add to the benchmark model county-specific coefficients on three aggregate time series: the price of oil, aggregate construction employment, and reserve balances with the Fed system. We show these estimates in Panels 2 and 3, respectively, of **Table B3**. The first column shows HMM's estimate and our replication using the pre-redesign LAUS data.[70] These estimates are qualitatively similar to the estimates from the benchmark model. And, like the benchmark model, the coefficient estimates come much closer to zero when post-redesign LAUS employment data is used, consistent with the null effect of benefit expansions that we find in our baseline specifications. We have not been able to replicate their results with the QCEW.

Additionally, HMM estimate "event study" specifications, as described in their Section 6. These specifications are designed to compare employment in 2014 to what is predicted to have occurred in the absence of the EUC expiration based on pre-expiration data. These predictions are formed by estimating a model using data solely from 2005q1 to 2013q4, and by using the resulting parameter estimates to project the future path of employment in a given county. To estimate the pre-event model, HMM regress county-level log employment on county fixed effects, date fixed effects, a county-specific cubic in the quarterly date, and four lags of log employment. They then define their dependent variable e^*_{ct} as the difference between actual log employment and predicted log employment based on the model parameters. Finally, they recover the effect of the EUC expiration by estimating the following model using observations only from 2014:

$$e^*_{ct} = \alpha\big(\ln(D_{c,2014}) - \ln(D_{c,2013q4})\big) + \nu_{pt} + \epsilon_{cpt} \tag{B3}$$

[70] We calculate standard errors in Panel 2 via a block bootstrap at the state-pair level. We use four factors, as HMM report using for LAUS employment, throughout Panel 2.

They estimate a coefficient of approximately -0.02, both using employment from LAUS and from the QCEW, meaning that counties which saw larger declines in benefits than their neighbors (i.e., whose independent variable is more negative) experienced higher growth of log employment in 2014, relative to their neighbors, relative to the prediction of their model. As with the estimates found in the "interaction term" models using pre-revision LAUS, this estimate would imply that the 26-to-99 week expansion would explain about one third of the EPOP drop during the Great Recession.

While we have not been able to replicate their results exactly, we do obtain qualitatively similar results. The main result from the event study strategy can be seen immediately in **Appendix Figure B2**, which plots the time series of the average value of log employment, as well as the series of predicted log employment, for high-benefit counties relative to low-benefit counties (where "high" and "low" status is defined by the size of the drop in log duration between 2013q4 and 2014, relative to the county pair partner). The model predicts that employment in high-benefit counties will continue to fall in 2014 relative to their lower-benefit neighbors, when in fact, a modest reversal occurs. The event study approach attributes this to the effect of the EUC expiration. As in the "interaction term" models discussed above, the redesign of the LAUS series affects the results substantially. When we repeat the analysis using the revised data, we find that the coefficient estimate becomes slightly (and insignificantly) positive, as shown in **Appendix Figure B3**. HMM also estimate the event study with QCEW data, and find an estimate of -0.0236, which is larger (in magnitude). When we estimate this model using employment from the QCEW, we find a coefficient of -0.0126 (with a standard error of 0.0113), which is in between our estimates for the specifications with revised and vintage LAUS log employment, respectively.[71] This is shown graphically in **Appendix Figure B4**.

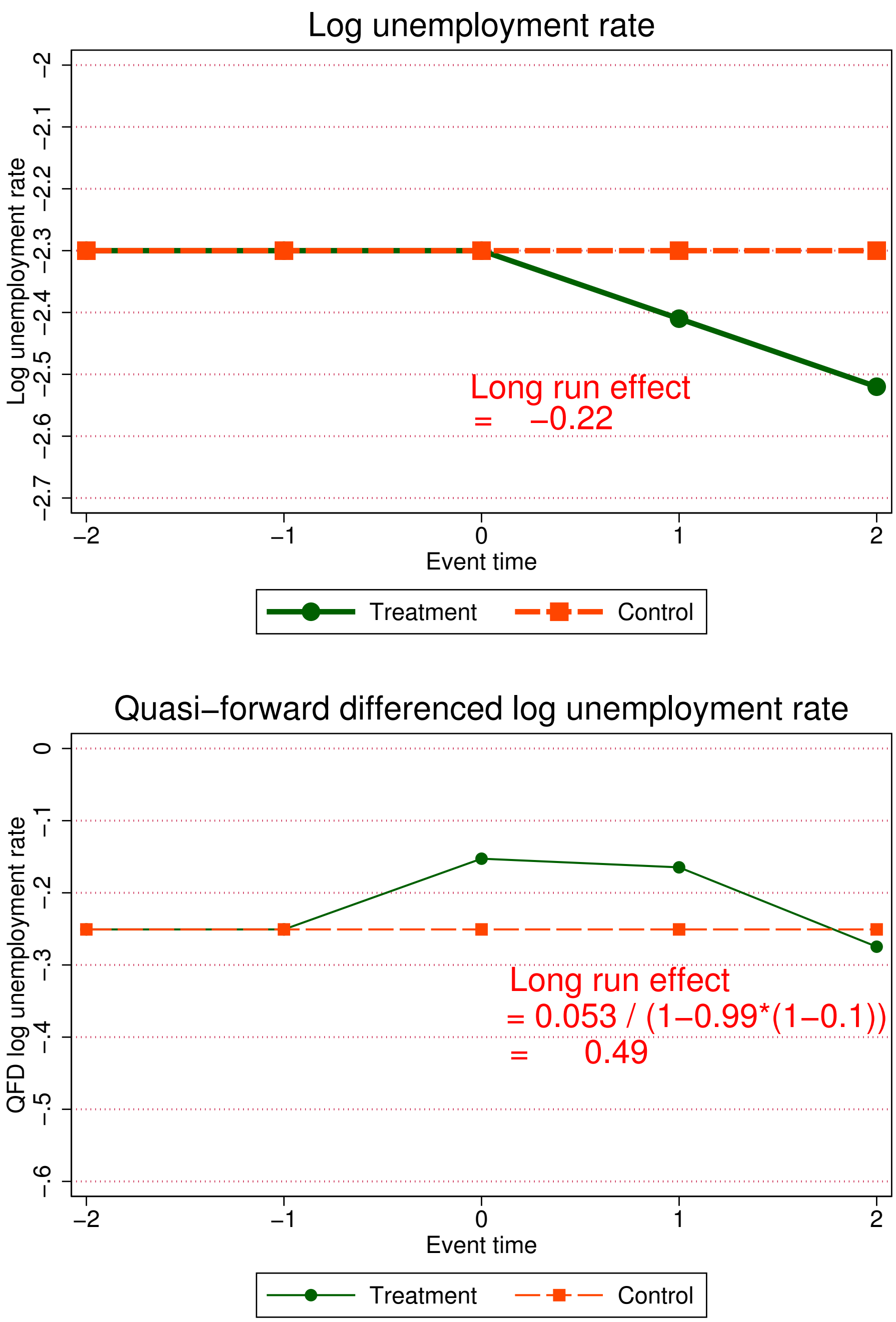
When translated to a change in EPOP, our replication of HMM's event study estimate using the QCEW (-0.703) is substantially more negative than our estimates using EUC expiration, which ranged between -0.024 (full BCP-FE sample) and -0.182 (PT-trimmed sample). HMM's event study strategy estimates a negative effect of EUC expiration using QCEW data because it constructs a counterfactual where the employment differential between the high and low treatment counties is expected to become more negative in 2014. This HMM counterfactual is largely driven by a county-specific polynomial time trend, whose identification is heavily reliant on employment changes that occur up to nine years before the treatment event.[72] As an indication of the type of problem with such a parametric strategy, the employment reversal (both in the QCEW data and, in fact, in the pre-revision LAUS data as well) appears to begin a few quarters prior to the expiration of EUC—a "pre-reversal" which casts doubt on the plausibility of a continuing downward trend as the appropriate counterfactual. In contrast, we take a much more flexible approach by showing whether

[71] This standard error takes the parameters of the model estimated in the pre-change period as non-random, likely causing us to understate this standard error. HMM use a bootstrapping procedure to construct these standard errors.

[72] The use of a cubic trend, rather than some other degree of polynomial, does not affect these results substantially.

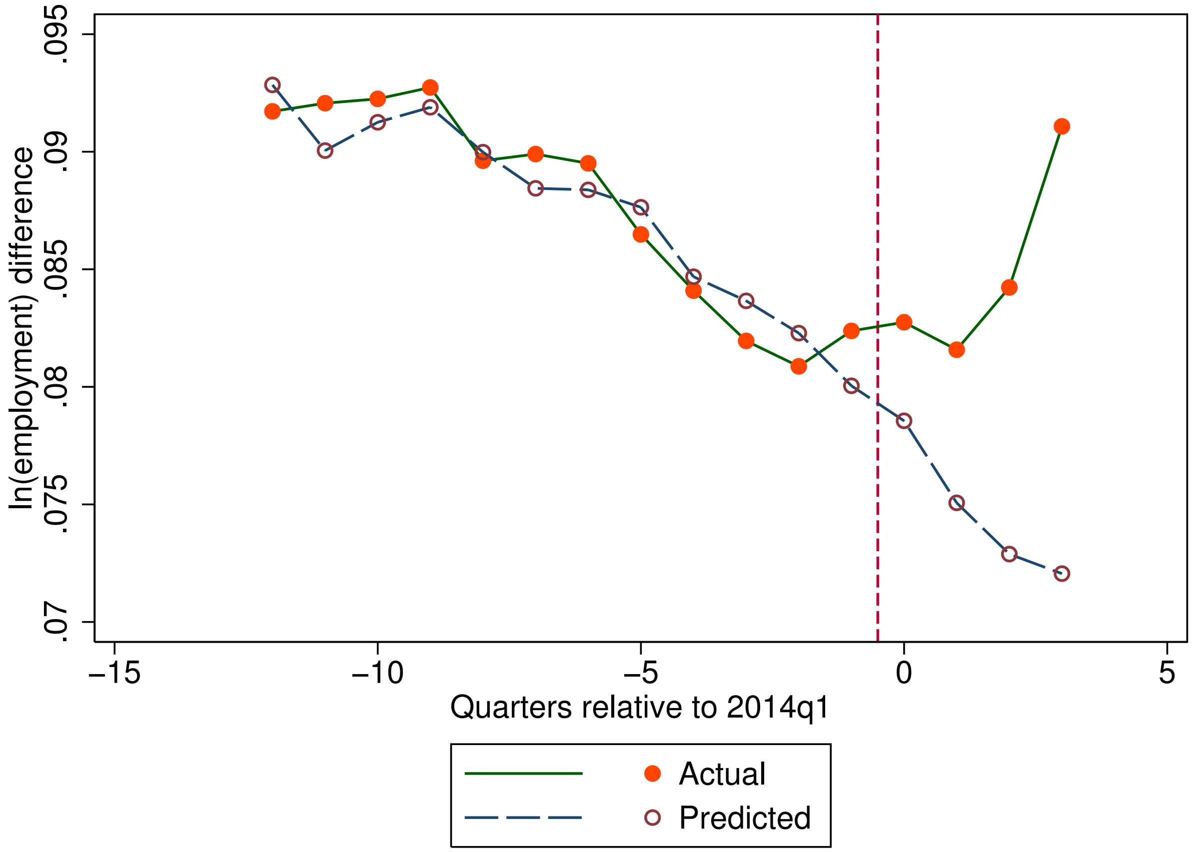
the employment rates were following parallel trends prior to 2014 by treatment status on the two sides of the border in our 2014 expiration IV. We find that they were, indeed, following parallel trends—as shown clearly in **Figure 7** for the full set of border county pairs. And that this employment gap between the two sides of the border remained largely unchanged following the 2014 expiration. We think the more transparent evidence from the 2014 event that we provide in **Figure 7** raises questions about the causal import of the parametric model used by HMM to construct the counterfactual employment path.

Figure B1: Illustration of wrong-signed estimate using quasi-forward-differenced data



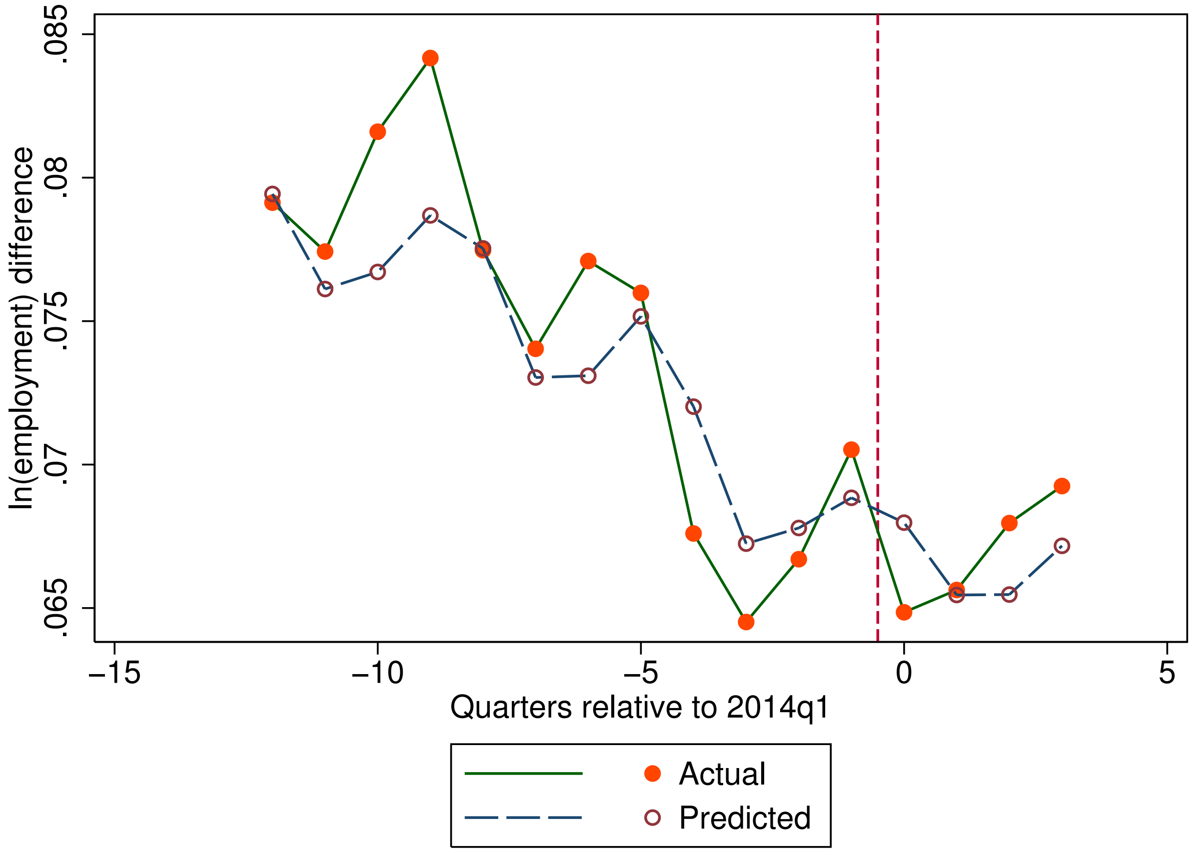
Notes: These two figures illustrate how using a quasi-forward-differenced dependent variable can lead to an estimate that is wrong-signed. In the top panel, the log unemployment rate of a hypothetical treated county is plotted against that of a control county. At event time zero, a one-unit increase in UI duration occurs. The control county unemployment rate is unchanged, but the treatment county log unemployment rate falls by 0.11 at event time 1 and an additional 0.11 at event time 2, where it remains at event time 3 (not plotted). The long-run effect of this change is a reduction of the unemployment rate by 22 log points. In the bottom panel, the data is plotted after taking a quasi-forward-difference (assuming a separation rate of 10 percent). The quasi-forward-differenced unemployment rate increases in the treated county and is unchanged in the control county. The implied coefficient estimate is 0.053, which erroneously implies an increase in the steady state unemployment rate of 49 log points – the wrong sign relative to the true effect.

Figure B2: Replication of HMM event study: Pre-revision LAUS employment



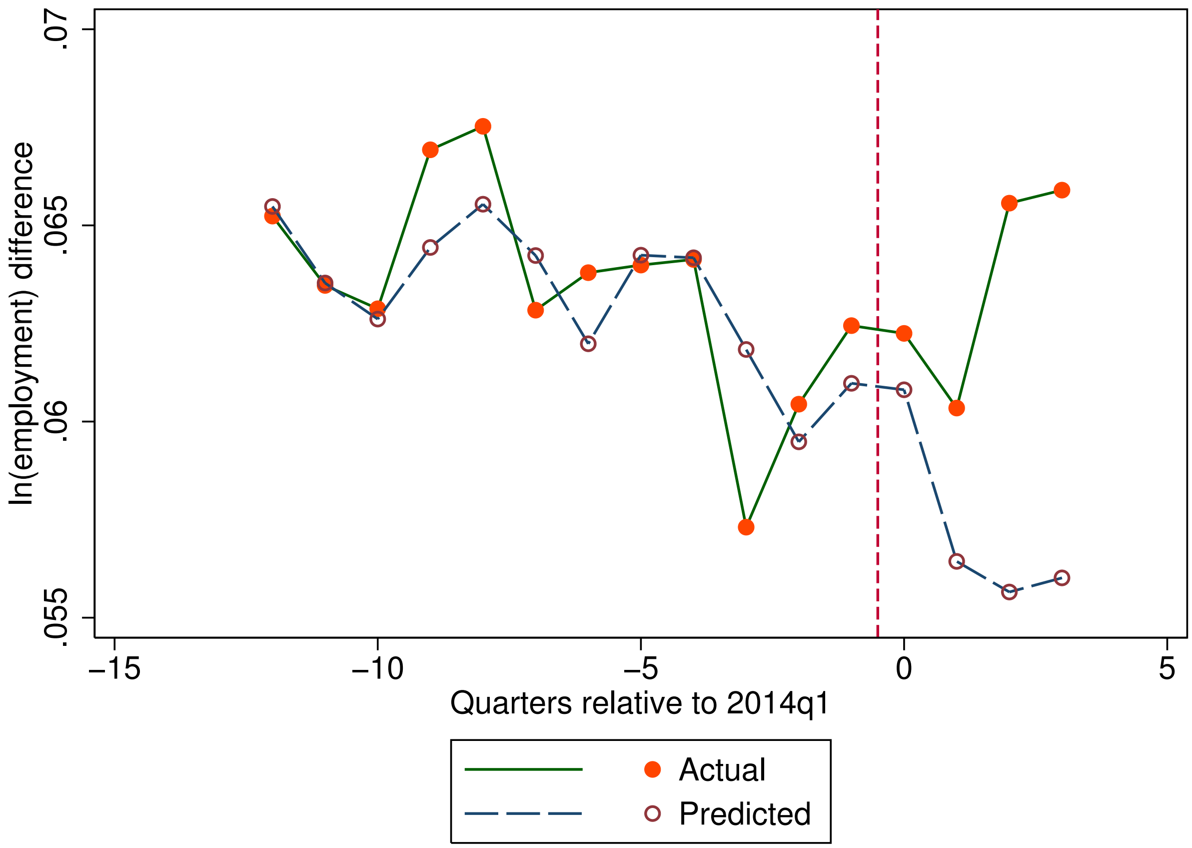
Notes: This figure plots (solid line, solid points) the average difference in log employment between "high" and "low" counties, where a "high" county is defined to have experienced a larger drop in log duration between 2013q4 and 2014 than its neighbor; pairs which experienced identical drops in log duration are not included. The figure also plots (dashed line, hollow points) the average difference in predicted log employment between high and low counties, where the prediction is computed by regressing (on quarterly data from 2005q1 through 2013q4) county log employment on four lags of log employment, time fixed effects, and a county-specific cubic function of the date. Predictions in 2014q1 through 2014q4 are computed recursively. This figure uses employment data from LAUS, prior to the March 2015 redesign.

Figure B3: Replication of HMM event study: Post-revision LAUS employment



Notes: This figure plots (solid line, solid points) the average difference in log employment between "high" and "low" counties, where a "high" county is defined to have experienced a larger drop in log duration between 2013q4 and 2014 than its neighbor; pairs which experienced identical drops in log duration are not included. The figure also plots (dashed line, hollow points) the average difference in predicted log employment between high and low counties, where the prediction is computed by regressing (on quarterly data from 2015q1 through 2013q4) county log employment on four lags of log employment, time fixed effects, and a county-specific cubic function of the date. Predictions in 2014q1 through 2014q4 are computed recursively. This figure uses current LAUS data.

Figure B4: Replication of HMM event study: QCEW employment



Notes: This figure plots (solid line, solid points) the average difference in log employment between "high" and "low" counties, where a "high" county is defined to have experienced a larger drop in log duration between 2013q4 and 2014 than its neighbor; pairs which experienced identical drops in log duration are not included. The figure also plots (dashed line, hollow points) the average difference in predicted log employment between high and low counties, where the prediction is computed by regressing (on quarterly data from 2015q1 through 2013q4) county log(employment) on four lags of log employment, time fixed effects, and a county-specific cubic function of the date. Predictions in 2014q1 through 2014q4 are computed recursively. This figure uses employment data from QCEW.

Table B1: Decomposition of difference between estimates from HKMM and BDGK into contributing factors

Step	From HKMM	To BDGK	Average Marginal Effect
Base Estimate	-2.6612	0.4299	
	(0.6298)	(0.4867)	
No QFD	2.6184		1.3016
	(0.5953)		(0.4197)
Align sample period	0.8165	1.5719	0.9149
	(0.6915)	(0.8585)	(0.3487)
Urate to EPOP	1.1196	1.1334	0.6731
	(1.1841)	(0.4775)	(0.3946)
Bai to FE	0.8241	0.6469	0.1891
	(0.7205)	(0.5576)	(0.2952)
Logs to levels		0.0777	0.0088
		(0.3403)	(0.1416)
Align counties	-0.1162	0.0979	0.0035
	(0.0846)	(0.0658)	(0.0267)

Notes: The first row reports the total effect of the expansion of UI from 26 to 99 weeks, in percentage points of EPOP, implied by the coefficient estimates of HKMM (column 1) and the full sample BCP-FE estimates of this paper (BDGK) (column 2). The remaining estimates in the first column represent the increased total implied effect of UI when one specification change is made from the original HKMM estimate. The remaining estimates in the second column represent the effect of taking each final step to arrive at the BDGK estimate. Because the total implied effect is not well motivated by theory when using quasi-differenced data in levels, we leave two cells blank in these first two columns. The third column represents the average incremental effect of taking each step along all possible transition paths between HKMM and BDGK estimates, except that we discard transition paths that involve estimating models with quasi-differenced data in levels. See text for details regarding each step and the conversion of each coefficient estimate into an effect on EPOP. Standard errors are calculated via a block bootstrap at the state-pair level with 200 replications.

Table B2: Transitioning from HKMM to BDGK estimates: Contribution of factors along three particular paths

	Path 1			Path 2			Path 3	
	Coefficient	EPOP effect		Coefficient	EPOP effect		Coefficient	EPOP effect
HKMM reported result	0.0490	-2.5885						
HKMM replication	0.0510	-2.6612	HKMM replication	0.0510	-2.6612	HKMM replication	0.0510	-2.6612
	(0.0097)	(0.6832)		(0.0097)	(0.6832)		(0.0097)	(0.6832)
Align counties	0.0527	-2.7774	Urate to EPOP	-0.0029	-1.5416	Align sample period	0.0149	-1.8217
	(0.0083)	(0.5905)		(0.0019)	(0.9967)		(0.0032)	(0.4955)
Eliminate QD	0.0104	-0.0428	Elimate QD	-0.0029	-0.1742	Align counties	0.0153	-1.8778
	(0.0332)	(0.1376)		(0.0055)	(0.3247)		(0.0029)	(0.4411)
Bai to FE	0.1291	-0.5759	Logs to levels	-0.0409	-0.0409	Elimate QD	0.0061	-0.0251
	(0.0428)	(0.2077)		(0.2700)	(0.2700)		(0.0203)	(0.0837)
Urate to EPOP	-0.0300	-1.7525	Align sample period	-0.2173	-0.2173	Logs to levels	0.3197	-0.2046
	(0.0125)	(0.7149)		(0.1642)	(0.1642)		(0.1595)	(0.1021)
Align sample period	0.0061	0.3641	Bai to FE	0.4351	0.4351	Bai to FE	1.0995	-0.7035
	(0.0083)	(0.4988)		(0.4802)	(0.4802)		(0.2498)	(0.1599)
Logs to levels (BDGK)	0.4299	0.4299	Align counties (BDGK)	0.4299	0.4299	Urate to EPOP (BDGK)	0.4299	0.4299
	(0.4711)	(0.4711)		(0.4711)	(0.4711)		(0.4711)	(0.4711)

Notes: This table presents three transition paths from HKMM's estimates to the full sample BCP-FE estimates of this paper (BDGK). Each cell presents the coefficient estimate, as well as the implied total effect of the 26-99 week expansion of UI expressed as an implied impact of EPOP, in percentage points. Once a step is made in a given path, it is retained in subsequent specifications in the same path. See text for details regarding each step. Standard errors for specifications involving the Bai (2009) interactive effects estimator are calculated via a block bootstrap at the state-pair level with 200 replications. Standard errors for other specifications are clustered twoway at the state and state-pair level.

Table B3: Estimates using the HMM interaction-term model: Alternative data sets and specifications

	(1) LAUS (orig.)	(2) LAUS (rev.)	(3) QCEW
Benchmark			
HMM's estimate	-0.0190		-0.0100
	[0.000]		[0.050]
Our estimate	-0.0200	-0.0048	-0.0078
	(0.0082)	(0.0060)	(0.0069)
Observations	46440	46440	46440
Interactive Effects			
HMM's estimate	-0.0233		-0.0121
	[0.000]		[0.030]
Our estimate	-0.0231	-0.0050	-0.0031
	(0.0093)	(0.0082)	(0.0092)
Observations	92720	92720	92880
Natural Factors			
HMM's estimate	-0.0144		-0.0141
	[0.000]		[0.020]
Our estimate	-0.0138	-0.0013	-0.0065
	(0.0104)	(0.0070)	(0.0067)
Observations	46440	46440	46440

Notes: This table reports estimates of α from HMM's "interaction-term" model: $e_{ct} = \kappa[\ln(D_{ct})\mathbb{1}(t \leq 2013q3)] + \alpha[\ln(D_{ct})\mathbb{1}(t \geq 2013q4)] + \nu_{pt} + \epsilon_{cpt}$, under different characterizations of the error term ϵ_{cpt}. In each panel, the top row reports the estimates reported by HMM, with p-values (from a block bootstrap at the state-pair level) in brackets. The second row reports our replication, with standard errors in parentheses. The first column uses log employment from LAUS, prior to the 2015 redesign. The second column uses post-redesign LAUS data, downloaded on September 9, 2016. The third column uses (not-seasonally-adjusted) log employment from the QCEW. The first panel represents the "benchmark" specification, in which $\epsilon_{cpt} = \mu_c + \gamma_c t + u_{cpt}$. The second panel replaces the fixed effects and county-specific trends with interactive effects (Bai (2009)): $\epsilon_{cpt} = \lambda_c' F_t + u_{cpt}$. The third panel adds to the benchmark specification county-specific coefficients on three national time series: the price of oil, employment in the construction industry, and reserve balances with the Fed system. Standard errors in the first and third panel are analytical, clustered at the state-pair level. Standard errors in the second panel are derived from a block bootstrap at the state-pair level.

C Online Appendix C: Additional Tables and Figures

Figure C1: Increase in UI benefit duration from the November 2008 expansion of EUC

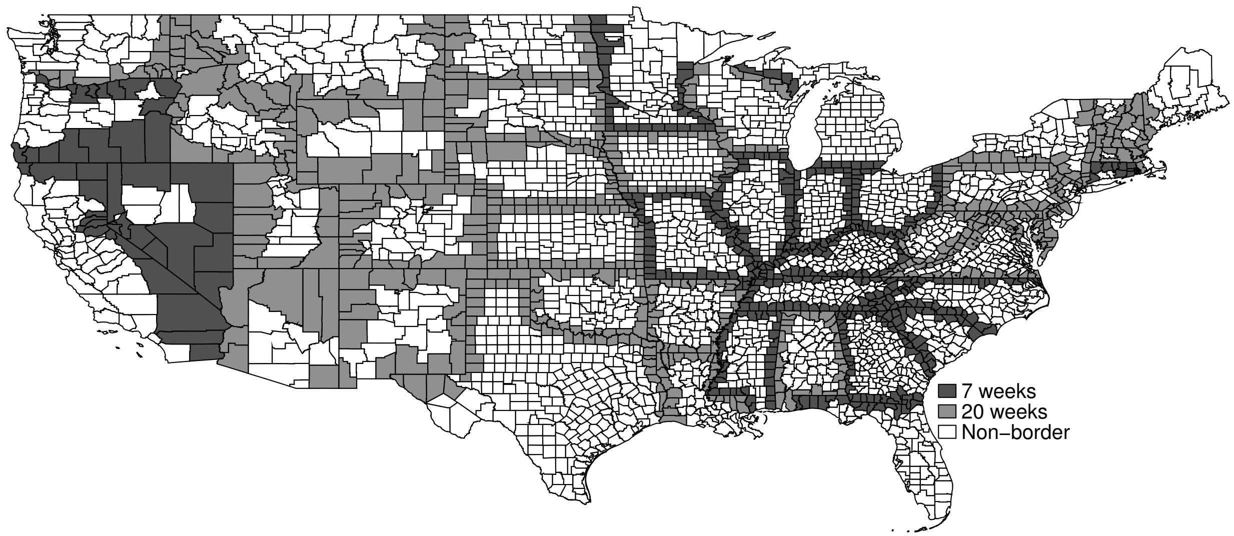
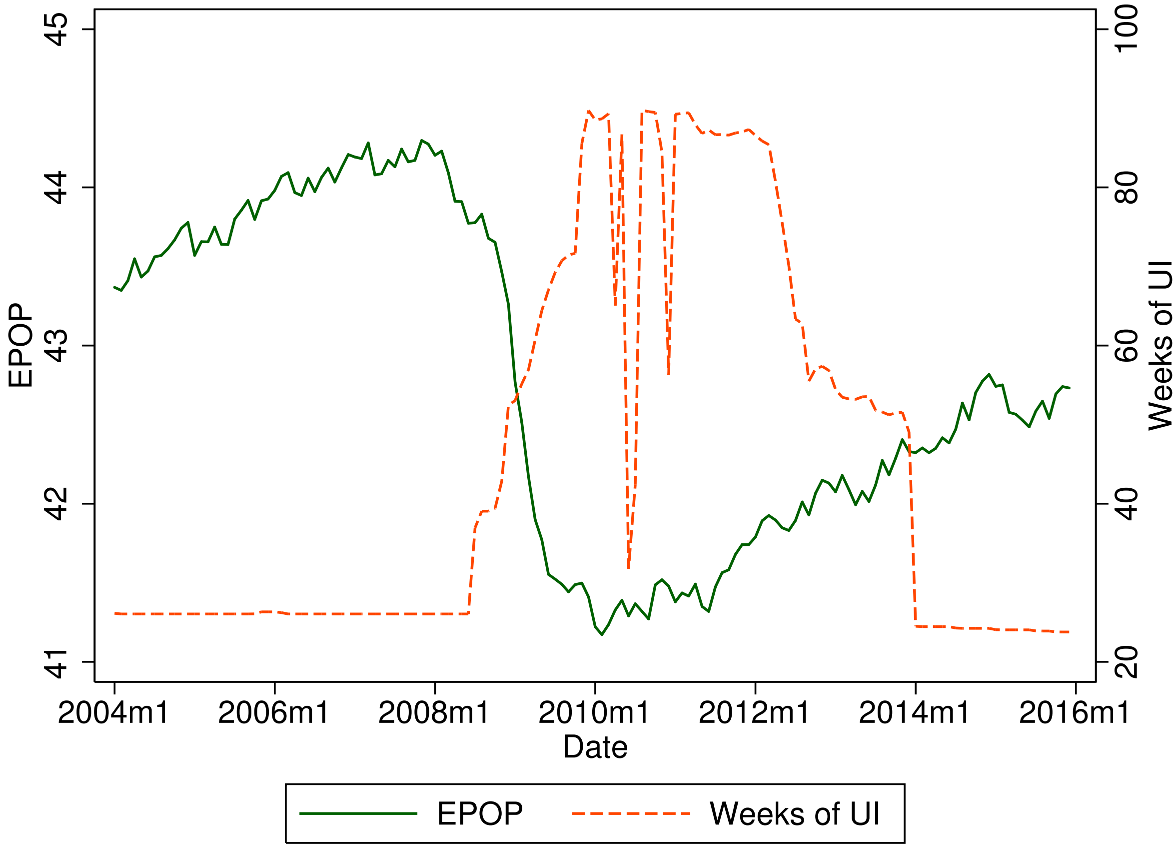
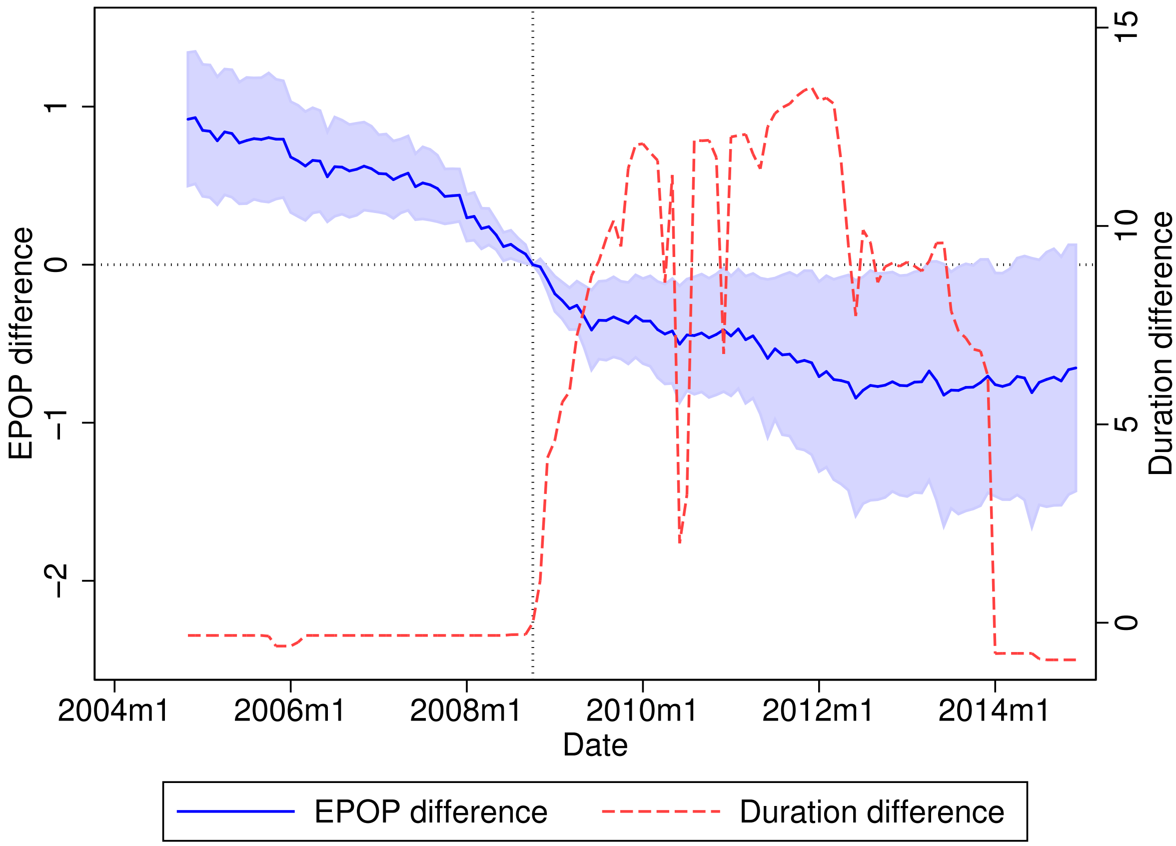


Figure C2: Evolution over time: national QCEW-based EPOP ratio and UI benefit duration



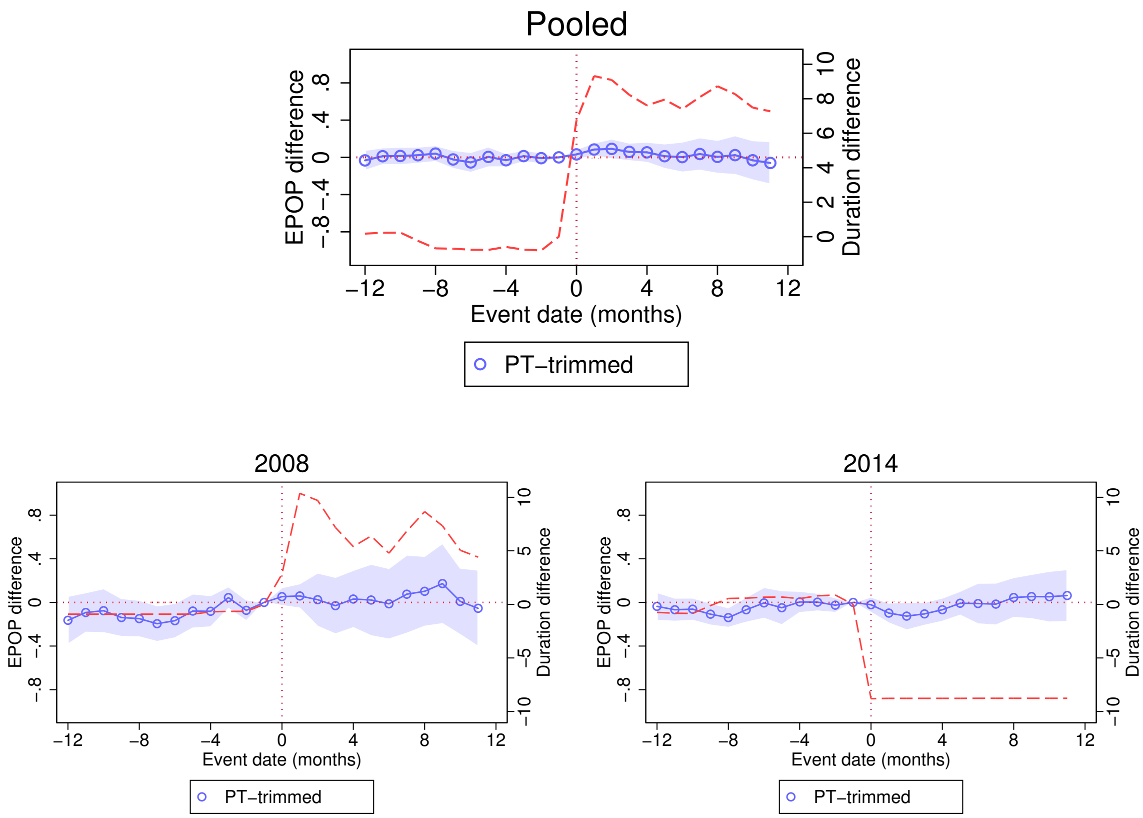
Notes: EPOP is the seasonally-adjusted ratio of employment (from the QCEW) to population age 15+. Weeks of UI represents the maximum number of weeks of UI compensation available. In this figure, both EPOP and weeks of benefits are calculated via an unweighted average of counties.

Figure C3: Evolution of EPOP and UI benefit duration differentials by average treatment intensity: without pair-period fixed effects, using all counties



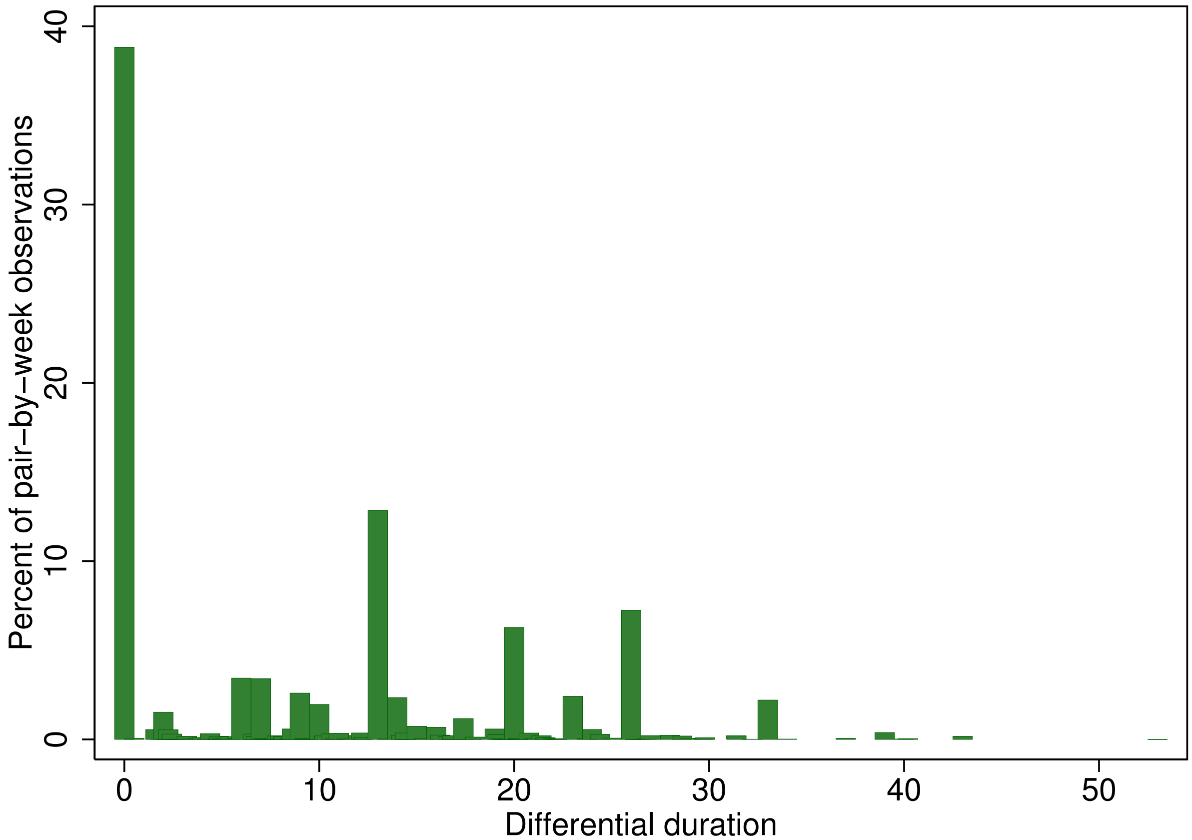
Notes: This figure plots (solid line, left axis) the set of β_s coefficients from the following regression: $E_{ct} = \sum_{s=\tau_A}^{\tau_B} \beta_s treat_c \mathbb{1}\{t = s\} + \lambda_c + \nu_t + \epsilon_{ct}$. E_{ct} is the seasonally-adjusted ratio of total employment to population age 15+, scaled in percentage points. The average treatment intensity, $treat_c$, is a time-invariant, continuous measure defined as the average duration during the treatment period (2008m11-2013m12), minus average duration from the 12 months prior (2007m11-2008m10), divided by 10. The shaded region corresponds to the 95% 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 2008m10, 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.

Figure C4: Evolution of EPOP difference and UI benefit duration difference across state borders: 2008 expansion and 2014 expiration of EUC, using PT-trimmed sample



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, using the TT-trimmed sample. The top panel uses pooled 2008 and 2014 samples, centered around event date -1 whose cumulative response is defined as zero. The bottom panel separately examines 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. 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).

Figure C5: Distribution of differences in UI benefit duration across border county pairs



Notes: This figure plots the distribution of duration differences across border county pairs, with each observation at the pair-by-(calendar)-week level. The sample is restricted to weeks between November 23, 2008, and December 22, 2013.

Figure C6: Distribution of EUC differences across border county pairs immediately prior to EUC expiration

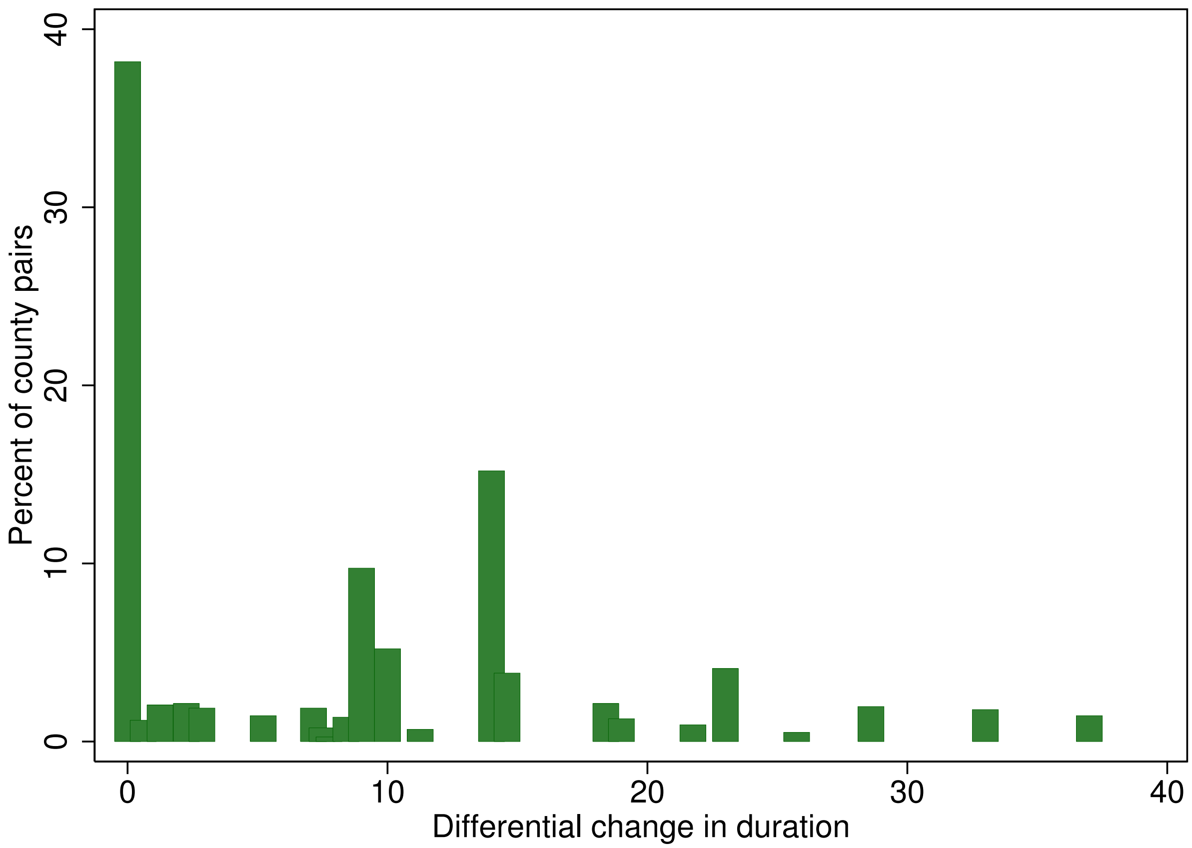
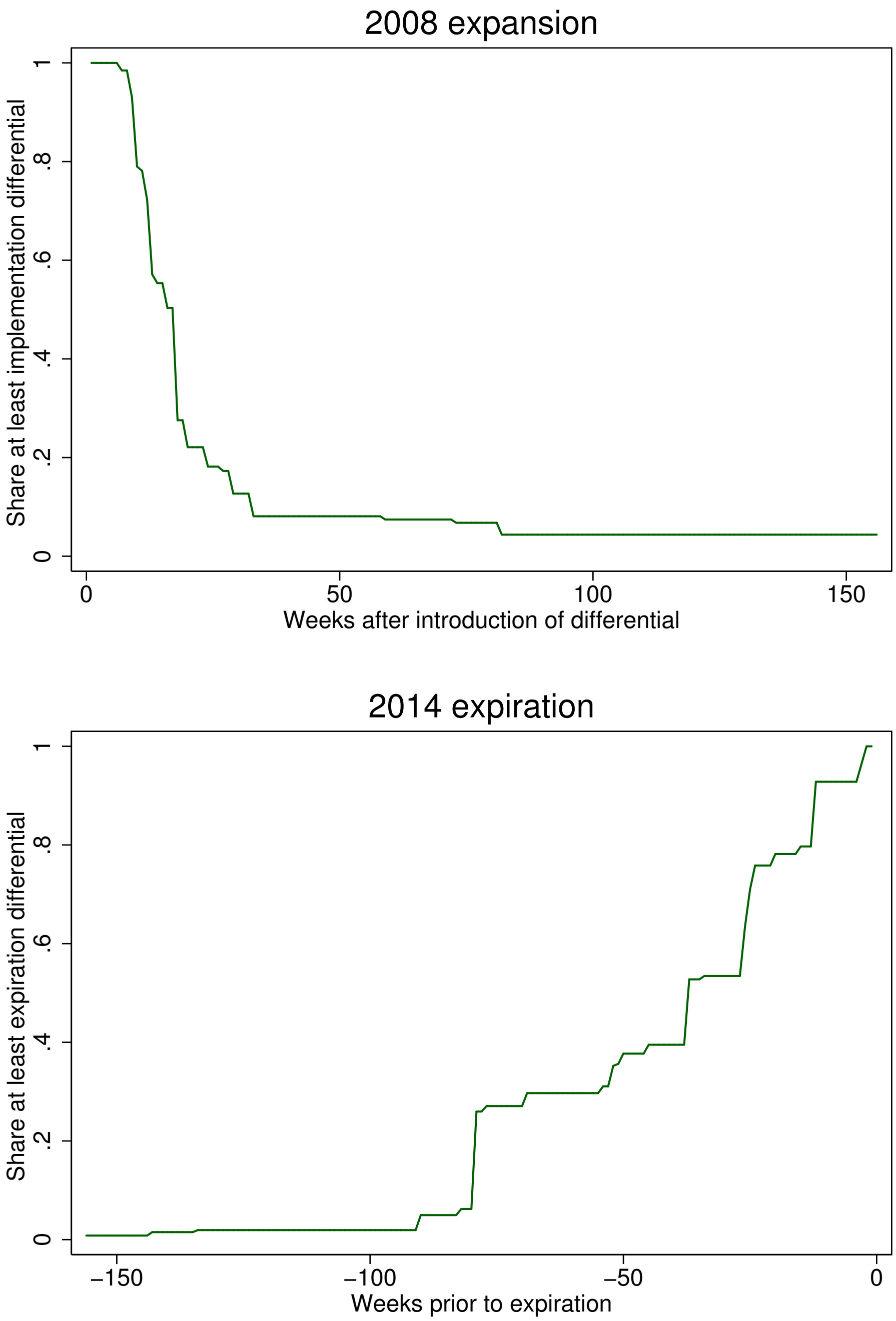


Figure C7: Persistence of duration differences in 2008 and 2014 events



Notes: The top graph plots the share of county pairs that continuously have a duration difference at least as large as immediately after the implementation of UCEA in November 2008. The bottom graph plots the share of county pairs that continuously have a duration difference (moving backward in time) at least as large as immediately prior to the 2014 expiration of EUC. The sample of pairs is restricted to those with differential duration at the time of the event in question.

Table C1: Summary statistics: High-treatment versus low-treatment counties, randomly-matched pairs versus border county pairs

	Randomly-matched				**Border pairs**			
	High: Mean	Sd	Low: Mean	Sd	High: Mean	Sd	Low: Mean	Sd
EPOP (A)	43.520	19.054	44.683	15.672	44.107	17.044	44.621	15.110
Private EPOP (A)	32.374	15.944	33.372	14.469	32.742	14.468	33.485	14.074
LAUS unemp. rate (A)	8.333	2.234	6.749	2.535	7.745	2.358	7.167	2.403
Population age 15+ (A)	100,302	225,744	56,185	133,920	81,415	213,875	71,441	153,060
Share white (B)	0.806	0.179	0.810	0.187	0.811	0.182	0.811	0.177
Share black (B)	0.090	0.145	0.089	0.154	0.085	0.145	0.086	0.147
Share hispanic (B)	0.069	0.105	0.057	0.100	0.067	0.111	0.059	0.092
Share H.S. grad (B)	0.565	0.066	0.571	0.063	0.569	0.064	0.567	0.065
Share college (B)	0.184	0.086	0.186	0.082	0.179	0.078	0.189	0.086
Median h.h. income (B)	43,557	11,842	42,450	11,293	42,645	11,459	43,535	12,127
New mortgage debt p.c. (A)	3.804	3.021	3.086	2.874	3.386	3.092	3.586	2.877
Share in cities 50k+ (C)	0.226	0.355	0.151	0.303	0.190	0.331	0.196	0.331
Min. weeks of UI elig.	24.032	3.876	24.122	3.591	24.470	3.495	24.631	3.199
Max. weeks of UI elig.	97.041	5.829	85.133	13.912	96.105	6.674	86.996	13.320
Pairs w/ different avg treatment	2317		2317		1131		1131	
Pairs w/ identical avg treatment	5		5		30		30	

Notes: The first four columns report summary statistics in county pairs, 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 2008m11-2013m12 period, minus average duration over the 2007m11-2008m10 and 2014m1-2014m12 periods. County pairs with identical treatment are dropped in this table. Columns 1-4 use a set of county pairs formed by randomly matching each county to some other county. Columns 5-8 report analogous statistics for the border county pairs. 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.

Table C2: Summary statistics for all counties, all county border pairs, and PT-trimmed sample of county border pairs

	All counties		**Border counties**		**PT-trimmed**	
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
EPOP (2007)	44.19	18.33	44.51	16.20	43.15	13.96
Private EPOP (2007)	34.58	17.45	34.88	15.47	33.94	13.35
LAUS unemployment rate (2007)	4.857	1.686	4.948	1.777	5.055	1.655
Population age 15+ (2007)	76818.0	243398.5	72692.4	178383.3	84309.4	193927.5
Share white (2005-2009 ACS)	0.796	0.190	0.812	0.181	0.814	0.179
Share black (2005-2009 ACS)	0.0885	0.144	0.0834	0.145	0.0872	0.144
Share hispanic (2005-2009 ACS)	0.0755	0.128	0.0620	0.101	0.0552	0.0922
Share high school grad, less than Bachelor's (2005-2009 ACS)	0.564	0.0665	0.568	0.0640	0.566	0.0635
Share Bachelor's degree or higher (2005-2009 ACS)	0.187	0.0852	0.184	0.0818	0.187	0.0838
Median household income (2005-2009 ACS), 2009 dollars	43299.6	11419.7	42949.1	11725.8	43258.3	12282.6
Newly acquired mortage debt per capita (2007)	3.535	3.216	3.508	3.120	3.604	3.036
Share in cities 50k+ (2010 census)	0.186	0.333	0.188	0.328	0.216	0.344
Minimum weeks of UI eligibility over sample period	23.78	4.365	24.17	4.040	24.26	3.931
Maximum weeks of UI eligibility over sample period	91.37	12.15	90.74	12.38	91.16	11.91

Notes: If a border county appears in j county-pairs in the sample in question, then it appears j times for the purpose of creating estimates in this table. 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).

Table C3: Pre-existing employment trends prior to November 2008 UI benefit expansion

	(1) All counties	(2) Border counties	(3) Border counties	(4) PT-trimmed
Treatment X Date	-0.780	-0.976	-0.241	-0.147
	(0.244)	(0.206)	(0.288)	(0.130)
Observations	148896	111456	111456	83520
County fixed effects	X	X	X	X
Pair-period fixed effects			X	X

Notes: In columns 1 and 2, each cell reports the coefficient on $treat_c \times t$ from a regression of the following form: $E_{ct} = \alpha \times treat_c \times t + \lambda_c + \theta_t + \epsilon_{ct}$. In columns 3 and 4, each cell reports the coefficient on $treat_c \times t$ from a regression of the following form: $E_{cpt} = \alpha \times treat_c \times t + \lambda_c + \nu_{pt} + \epsilon_{cpt}$. In all columns, the dependent variable is the seasonally-adjusted ratio of total employment to population age 15+, scaled in percentage points. The regression is estimated over the period 2004m11-2008m10 and t is time measured in months since 2014m11 divided by 48 (representing the 48 month period between the beginning and the end of this sample). The time-invariant variable $treat_c$ is the average treatment intensity for each county, defined as the average duration over the 2008m11-2013m12 period, minus average duration from the 12 months prior (2007m11-2008m10), divided by 10. Thus we interpret the coefficientα as the change in EPOP between 2004m11 and 2008m10 associated with an additional 10 weeks average higher UI duration between 2009m11 and 2014m1. In column 1, standard errors are clustered at the state level. In columns 2, 3, and 4, standard errors are clustered two-way at the state and state-pair level. Columns 4 report the estimates from the set 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).

Table C4: Cumulative response of EPOP from distributed lags specification: full sample regressions

	(1) BCP-FE		(2) PT-Trimmed	
	Leads	Lags	Leads	Lags
Contemp.		-0.006		-0.012
		(0.097)		(0.072)
Lead/lag 1	0	-0.123	0	-0.055
	(0)	(0.129)	(0)	(0.098)
Lead/lag 2	0.118	0.004	-0.020	-0.005
	(0.114)	(0.150)	(0.089)	(0.115)
Lead/lag 3	0.208	0.218	0.092	0.036
	(0.157)	(0.221)	(0.110)	(0.151)
Lead/lag 4	0.263	0.154	0.044	0.002
	(0.198)	(0.185)	(0.146)	(0.149)
Lead/lag 5	0.148	0.220	0.062	0.054
	(0.243)	(0.229)	(0.192)	(0.177)
Lead/lag 6	0.243	0.069	0.077	0.065
	(0.268)	(0.289)	(0.197)	(0.198)
Lead/lag 7	-0.030	0.239	-0.083	0.083
	(0.280)	(0.287)	(0.178)	(0.219)
Lead/lag 8	0.058	0.113	0.043	0.044
	(0.317)	(0.318)	(0.192)	(0.225)
Lead/lag 9	0.056	0.229	0.021	0.025
	(0.322)	(0.337)	(0.202)	(0.253)
Lead/lag 10	0.307	0.165	0.100	0.046
	(0.332)	(0.376)	(0.228)	(0.267)
Lead/lag 11	0.403	0.112	0.132	0.056
	(0.345)	(0.376)	(0.224)	(0.290)
Lead/lag 12	0.228	0.168	0.059	0.050
	(0.338)	(0.394)	(0.231)	(0.304)
Lead/lag 13		0.154		0.107
		(0.403)		(0.322)
Lead/lag 14		0.094		-0.014
		(0.410)		(0.317)
Lead/lag 15		0.193		-0.013
		(0.470)		(0.326)
Lead/lag 16		0.341		-0.010
		(0.445)		(0.304)
Lead/lag 17		0.273		0.021
		(0.468)		(0.309)
Lead/lag 18		-0.003		-0.167
		(0.499)		(0.305)
Lead/lag 19		0.155		-0.068
		(0.510)		(0.323)
Lead/lag 20		0.162		-0.030
		(0.548)		(0.341)
Lead/lag 21		0.168		-0.096
		(0.580)		(0.341)
Lead/lag 22		0.210		0.022
		(0.597)		(0.340)
Lead/lag 23		0.181		-0.075
		(0.583)		(0.351)
Lead/lag 24		0.340		-0.087
		(0.596)		(0.361)

Notes: This table reports cumulative monthly lags and leads estimated on the full sample (2007m11-2014m12), using all border county pairs (BCPs) (column 1) and the subset of BCPs in the PT-trimmed sample (column 2), 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 and is estimated using EPOP data from 2007m11-2014m12 (and thus duration data from 2005m11-2015m11) in first differences. The zeroth cumulative lag is equal to the estimated coefficient on contemporaneous duration. The jth cumulative lag is equal to the estimated coefficient on contemporaneous duration plus the sum of the estimated coefficient on the 1st through jth lag term. The jth cumulative lead is equal to the sum of the estimated coefficient on the 1st through the $j-1^{th}$ lead term. The 1st cumulative lead is normalized to zero. 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). Standard errors are clustered two-way at the state and state-pair level.

Table C5: Robustness of the effects of UI benefit duration on EPOP: choice of cutoffs for trimming on match quality

	(1) Full	(2) Pooled	(3) 2008	(4) 2014
Baseline	0.430	0.143	0.549	-0.024
	(0.471)	(0.974)	(2.541)	(0.568)
	N = 199692	N = 108000	N = 55728	N = 52272
10th percentile	0.132	0.269	0.844	0.023
	(0.291)	(0.592)	(1.252)	(0.454)
	N = 179568	N = 97056	N = 50112	N = 46944
20th percentile	0.121	0.235	1.182	-0.173
	(0.273)	(0.653)	(1.276)	(0.513)
	N = 159616	N = 86352	N = 44544	N = 41808
25th percentile	0.180	0.253	1.344	-0.214
	(0.268)	(0.650)	(1.253)	(0.523)
	N = 149640	N = 80928	N = 41760	N = 39168
30th percentile	0.217	0.018	0.947	-0.385
	(0.269)	(0.662)	(1.190)	(0.568)
	N = 139664	N = 75456	N = 38976	N = 36480
40th percentile	0.358	-0.007	0.892	-0.388
	(0.303)	(0.635)	(1.151)	(0.576)
	N = 119712	N = 64608	N = 33408	N = 31200
50th percentile	0.240	-0.201	-0.090	-0.247
	(0.315)	(0.542)	(1.194)	(0.456)
	N = 99760	N = 53952	N = 27840	N = 26112

Notes: Each cell reports the baseline coefficient from the full sample, pooled event sample, and 2008 and 2014 subsamples, estimated over a different subsample of border county pairs. The cells in row 1 correspond to the estimates in column 1 of Table 2. In the other rows, the sample of border county pairs (BCPs) is trimmed. First, we calculate the pair-specific mean squared error in EPOP between 2004m11-2008m10 (after partialling out a fixed level difference). We then rank and trim all BCPs according to these MSEs. In the second row, we drop the bottom 10 percent of BCPs with the largest MSE, in the third row, we drop the bottom 20 percent, and so forth. The fourth row (the 25th percentile) corresponds to the estimates in column 2 of Table 2. Standard errors are clustered two-way at the state and state-pair level.

Table C6: Additional robustness checks on the effects of UI benefit duration on EPOP

	2008 sample IV		2014 sample IV	
	(1) BCP-FE	(2) PT-Trimmed	(3) BCP-FE	(4) PT-Trimmed
1. Baseline	0.549	1.344	-0.024	-0.214
	(2.541)	(1.253)	(0.568)	(0.523)
2. Private EPOP	1.097	1.904	-0.164	-0.264
	(2.540)	(1.203)	(0.639)	(0.523)
3. Correlation-trimmed	0.559	1.163	-0.392	-0.327
	(2.895)	(1.198)	(0.728)	(0.581)
4. ISLT	0.237	0.578	1.206	0.830
	(1.398)	(0.715)	(0.878)	(0.392)
5. Quarterly data	0.787	1.560		
	(2.428)	(1.206)		
6. QWI EPOP (quarterly)	0.110	1.875	0.517	0.015
	(1.697)	(1.406)	(0.586)	(0.583)
7. Unbalanced panel	0.511	1.344	-0.002	-0.214
	(2.497)	(1.253)	(0.564)	(0.523)
8. ln($EPOP$)	0.032	0.018	-0.001	-0.008
	(0.045)	(0.025)	(0.009)	(0.010)
	[1.886]	[1.048]	[-0.037]	[-0.442]
9. ln(emp)	0.039	0.019	0.006	-0.001
	(0.045)	(0.026)	(0.010)	(0.010)
	[2.327]	[1.138]	[0.320]	[-0.046]
10. Exploit Δ reg. benefits			0.037	-0.194
			(0.561)	(0.502)
11. Drop NC	0.660	1.542		
	(2.838)	(1.388)		
12. Keep NC			-0.437	-0.667
			(0.717)	(0.724)
13. NC: Alt. instrument			-0.037	-0.178
			(0.326)	(0.296)
14. Distance trimming	1.482	2.250	-0.229	-0.354
	(2.592)	(1.126)	(0.712)	(0.640)
15. Hinterland pairs	-0.979	0.600	0.556	-0.800
	(1.367)	(1.532)	(1.726)	(0.671)

Notes: Each cell reports regressions analogous to those reported in Table 2 for the 2008 and 2014 estimates (IV), respcetively. The estimates in the 1st row correspond to the estimates in the top two panels of Table 2. The estimates in the 2nd row replace (total) EPOP with the ratio of private employment to population age 15+. In the 3rd row, we trim the set of border county pairs based on the level of correlation between county EPOP and state EPOP over the period 2004m11-2008m10 (see text for details). The 4th row controls for county-specific linear trends. The 5th row uses quarterly data instead of monthly (and estimates over the 2007q4-2014q4 period). The 6th row uses EPOP derived from the QWI (at the quarterly level) instead of the QCEW. The 7th row includes counties without full EPOP data for each month, which we drop by default. The 8th and 9th row 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 $(\frac{99}{26}^{\hat{\beta}} - 1)\bar{E}$, where $\bar{E}$ is the mean EPOP level in the given sample. The 10th row uses a modified version of the instrument z_{ct} which exploits all changes in benefits, including changes in regular benefits, which occur at the end of December 2013. Rows 11-13 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 dropped in the 2014 subsample. The 11th row completely drops all NC BCPs. The 12th row keeps all North Carolina BCPs. The 13th row keeps NC BCPs but redefines the instrument for NC counties (see text for details). The 14th row drops county-pairs whose population centroids are greater than 100km apart. The 15th row uses the "hinterland" pairs rather than the border pairs (see text for details). Cells which are not applicable in the given sample, or which provide estimates that are mechanically equal to the baseline estimates, are left blank. Standard errors are clustered two-way at the state and state-pair level.

Table C7: Robustness to inclusion of regression discontinuity controls

	No county fixed effects		County fixed effects	
	(1)	(2)	(3)	(4)
Border pairs	0.547	-27.247	0.303	0.612
	(15.454)	(35.819)	(0.300)	(1.155)
Add Hinterland pairs	-3.965	-26.000	0.335	0.553
	(10.123)	(19.193)	(0.277)	(0.659)
Pair-period effects	X	X	X	X
County fixed effects			X	X
Control for distance to border		X		X

Notes: This table reports results from regressions of EPOP on duration, with differing controls. In all specifications, state-pair by period fixed effects are included. In columns 1 and 3, we control for distance to border, with coefficients allowed to vary at the state by state-pair by period level. In columns 3 and 4, we add county fixed effects. To be consistent with Dieterle (2016), we weight the regressions by 2010 population, and do not allow repeated county-month observations (e.g., in cases when a county is a member of more than one pair). In the first row, we use only those counties that are members of a border pair. In the second row, we add counties that are members of a hinterland pair, as discussed in the text. Standard errors are clustered two-way at the state and state-pair level.

Table C8: Robustness to auxiliary specification differences in Dieterle et al. (2018)

	No county fixed effects		**County fixed effects**	
		Add RD controls		Add RD controls
Level-on-level				
Baseline	-3.965	-26.000	0.335	0.553
	(10.123)	(19.193)	(0.277)	(0.659)
Drop Hinterland	0.547	-27.247	0.303	0.612
	(15.454)	(35.819)	(0.300)	(1.155)
Unweighted	-8.721	-9.896	0.600	-1.757
	(4.960)	(15.950)	(0.495)	(1.150)
Original sample	-8.560	-8.870	0.329	-1.398
	(4.436)	(12.052)	(0.474)	(0.977)
Log-on-log				
Baseline	-3.777	-9.940	0.291	-0.142
	(4.773)	(6.769)	(0.177)	(0.419)
Drop Hinterland	-0.459	-7.716	0.122	-0.077
	(7.326)	(10.869)	(0.240)	(0.617)
Unweighted	-8.045	-10.957	0.557	-1.149
	(3.851)	(11.300)	(0.440)	(1.125)
Original sample	-8.570	-10.474	0.278	-0.807
	(3.409)	(9.078)	(0.470)	(1.091)
Pair-period effects	X	X	X	X
County fixed effects			X	X
Control for distance to border		X		X

Notes: This table reports results analogous to **Appendix Table C7**, with additional variations. Rows 1 and 2 of the top panel are identical to rows 2 and 1 in **Appendix Table C7**, respectively. The remainder of the rows modify the specification such that the specification becomes closer to the baseline regression in the top panel of **Table 2**. In the third row, the regression is not weighted by population. In the fourth row, we allow for repeated observations at the county-by-month level (e.g., when a county is in multiple pairs). The bottom panel is analogous to the top panel, except that it regresses log(EPOP) on log(duration); the coefficient estimates and standard errors are translated to the implied level effect at the mean. Standard errors are clustered two-way at the state and state-pair level.

Table C9: Effect of EPOP on duration, interacted with the centroid distance between county pairs

	Border pairs	Hinterland pairs	All pairs
1. D	0.634	1.918	0.634
	(0.548)	(1.236)	(0.548)
	[163]	[378]	[163]
2. $D \times 1(30km \leq dist \leq 40km)$	-0.217		-0.242
	(0.805)		(0.802)
	[222]		[223]
3. $D \times 1(40km \leq dist \leq 50km)$	0.104		0.444
	(0.714)		(0.760)
	[259]		[260]
4. $D \times 1(50km \leq dist \leq 70km)$	-0.930		-0.252
	(0.674)		(0.988)
	[299]		[317]
5. $D \times 1(70km \leq dist \leq 90km)$	-0.142		-0.472
	(1.861)		(1.039)
	[228]		[189]
6. $D \times 1(90km \leq dist \leq 110km)$			0.654
			(0.521)
			[295]
7. $D \times 1(110km \leq dist \leq 135km)$		-0.831	0.607
		(1.231)	(0.535)
		[514]	[549]
8. $D \times 1(135km \leq dist \leq 160km)$		-1.245	-0.070
		(1.328)	(0.739)
		[331]	[350]
9. $D \times 1(160km \leq dist \leq 200km)$		-2.002	-0.497
		(1.481)	(0.925)
		[224]	[241]
10. $D \times 1(200km \leq dist)$		-1.430	-0.193
		(2.321)	(1.838)
		[386]	[417]
P-value for test of joint significance for interaction terms	0.644	0.625	0.576

Notes: This table reports coefficient estimates from a regression of EPOP on duration interacted with the centroid distance between pairs, in bins. The regression uses the 2007m11-2014m12 period and includes county fixed effects and pair-period fixed effects. The first column uses only border pairs. The second column uses only hinterland pairs, as discussed in the text. The third column pools both sets of pairs together. Standard errors are clustered two-way at the state and state-pair level. The number of pairs within each bin is reported in brackets below the estimate. Note that bin 5 in column 1 aggregates bins 5 through 10; likewise, bin 1 in column 2 aggregates bins 1 through 6.

D Online Appendix D: Joint Replication of HKMM

In this joint appendix, we describe our replication of the main estimate in Marcus Hagedorn, Fatih Karahan, Iourii Manovskii and Kurt Mitman (2015), hereafter HKMM. We face two challenges in the replication: (1.) properly constructing the data set and (2.) executing the estimation correctly. Though proper execution of the estimation is potentially challenging due to the non-linearity of the model, replicating the data set is ultimately more difficult. While we are unable to simultaneously reproduce the exact sample size and point estimate, our preferred replication is very close: we come within 0.002 of the estimate and our data set contains about 1% fewer observations (385 out of 37,177). In this description of our replication of HKMM, we discuss not only our final specification but also the choices we made and the reasons for those choices. In particular, we traded off sample size with estimate closeness. We considered specifications with closer estimates where the gap in the sample size compared to HKMM's sample was substantially larger; we also considered samples which matched more closely the sample size of HKMM but where the estimates were not as close.

Estimation Equation and Method

The HKMM estimation equation, which uses data at the pair-by-time level, is as follows. Prior to estimation, the data for a given pair p at time t is spatially differenced.

$$ln(u_{pt}) - \beta(1 - s_t)ln(u_{pt+1}) = \alpha * ln(D_{pt}) + \lambda_p' F_t + \epsilon_{pt}$$

In this expression,u_{pt} is the unemployment rate from LAUS (as calculated prior to the March 2015 redesign of the LAUS program), β is the discount factor equal to 0.99, s_t is the separation rate, andD_{pt} represents weeks of UI benefits available. $\lambda_p' F_t$ represent the interactive effects: F_t is a time-specific vector of length K of common factors, while λ_p (also of length K) represents the pair-specific factor loadings. HKMM determine, by minimizing an Akaike information criterion, that the optimal K is equal to 2. We replicate their minimization, also obtaining two factors as optimal. All of our estimates estimate with two factors in both space and time. We follow HKMM in estimating the model using the method of Bai (2009). In the April 2016 version of HKMM, the authors report a main estimate of 0.049.

Sample

The biggest challenge in replicating this result is determining precisely which pair-time observations were used in the sample. HKMM report using an unbalanced panel of quarterly LAUS unemployment data spanning 32 quarters from 2005q1 to 2012q4 with a sample size of 37,177 county-pair-by-quarter observations in their baseline regression. Dividing the number of observations by the number of quarters indicates that this sample size is similar to a balanced panel of 1,162 county-pairs (37,177/32=1,161.78). Our initial sample of pre-revision LAUS data yields an unbalanced panel for 1,171 county-pairs and a total number of observations of 37,464. This is a nearly balanced panel. It only drops data for the four quarters following Hurricane Katrina (2005q3-2006q2) for the two border pairs that include St. Tammany, LA (paired with Hancock, MS and Pearl River, MS). The missing counties for these quarters range from small to above mean county size. In 2005, these three counties had populations of 217,407, 46,097, and 51,764 respectively, according to the U.S. Census Bureau. Dropping these counties fully to create a balanced panel is essentially inconsequential to our estimates. The estimate in the balanced panel is 0.0527 and the estimate in the unbalanced panel is 0.0529.[73]

Though estimating on the above unbalanced sample yields estimates which are close to those reported in HKMM, in order to more closely match HKMM's reported number of observations, we consider an additional sample restriction to the unbalanced panel. In particular, we note that HKMM draw employment data for auxiliary specifications from the Quarterly Workforce Indicators (QWI). In an earlier draft from October 2013, HKMM report a sample size of 30,988 county-pair-by-quarter observations covering the period from 2005q1 to the beginning of 2012. Over the 28 quarters covered, this sample size would be consistent with a nearly balanced panel of 1,107 county pairs (30,988/28=1,106.71), which aligns with the number of pairs for which HKMM report having "complete data" in the October 2013 and April 2016 drafts. We believe that the phrase "complete data" likely refers to the presence of unemployment data in the LAUS and employment data in the QWI in a given quarter.

So, we consider the possibility that the choice to use the panel of "complete data" in the October 2013 draft may have carried over in some form to the sample used in the April 2016 draft. Specifically, since QWI data only cover Massachusetts beginning in 2010q1,[74] we exclude county pairs that include a Massachusetts county from the sample to generate an unbalanced panel of 1,150 county pairs. This sample restriction leads to a sample size of 36,792 county-pair-by-quarter observations. While this restriction leads to a smaller sample than reported in the April 2016 version of HKMM, it is simpler and yields an estimate closer to HKMM than other QWI-based sample restrictions we considered. In particular, using this sample, we find an estimate of 0.0510 (to four deminal places), compared to the 0.049 (to three decimal places) reported by HKMM.[75]We use this sample for the replication estimate used both by Dieterle, Bartalotti and Brummet

[73]See https://www.bls.gov/katrina/lausquestions.htm for a discussion of the impact of Hurricane Katrina on LAUS.

[74]The QWI includes both beginning-of-quarter and end-of-quarter statistics. Since the beginning-of-quarter statistics are rolled over from the previous quarter's end-of-quarter numbers, some of the QWI data for Massachusetts does not begin until 2010q2.

[75]We estimate this model using the user-written Stata command "regife" (Gomez 2015). We have also written our own simplified version of this command and are able to obtain identical estimates.

(2020) and Boone et al. (2018): an unbalanced panel which (1.) keeps counties which temporarily did not report in the aftermath of Hurricane Katrina and (2.) drops all counties in Massachussetts.

Possible Reasons for Remaining Discrepancy

Lastly, we note that there may be other minor specification choices that prevent us from replicating the results of HKMM. First, we obtain the dependent variable (unemployment rates as estimated by LAUS prior to the March 2015 redesign) through a FRED API. While the original LAUS dataset (which HKMM presumably used) includes the estimates of the raw counts of unemployed persons and the size of the labor force, the FRED API reports only the unemployment rate to the nearest tenth of a percentage point. Thus, HKMM may have been using unrounded unemployment rates while we are using rounded unemployment rates. Second, there may be differences in how we aggregate weeks of benefits from the weekly level (at which they are reported) to quarters. We calculate the weeks of benefits available on a given calendar day, and then aggregate to the month level. We then aggregate to the quarterly level using an unweighted average of the three months within the quarter. It is possible that HKMM performed this aggregation somewhat differently. Third, it is possible that we use a different separation rate than HKMM. We use the non-seasonally-adjusted total separation rate as reported by JOLTS. Other possibilities include the seasonally-adjusted version or the version which includes only private employment. In any case, while these uncertainties might prevent us from replicating HKMM's result exactly, the fact that our replication is within 0.002 (to three significant digits) of HKMM's estimate suggests that these minor differences do not matter qualitatively. [76]

[76]Since the Bai (2009) estimator is non-linear, an additional possibility is that the likelihood function used in the optimization has multiple local optima and that HKMM and our replication of HKMM are at different optima. We do not, however, think this is likely given (1.) that we are able to exactly replicate the optimality of two factors and (2.) that our estimates are so close to those of HKMM.