Working Papers

FEDERAL RESERVE BANK OF BOSTON™

No. 22-15

The Impact of the Age Distribution on Unemployment: Evidence from US States

Bruce Fallick and Christopher L. Foote

Abstract:
Economists have studied the potential effects of shifts in the age distribution on the unemployment rate for more than 50 years. Most of this analysis uses a "shift-share" method, which assumes that the demographic structure has no indirect effects on age-specific unemployment rates. This paper uses state-level data to revisit the influence of the age distribution on unemployment in the United States. We examine demographic effects across the entire age distribution rather than just the youth share of the population—the focus of most previous work—and extend the date range of analysis beyond that which was available for previous research. We find that shifts in the age distribution move the unemployment rate in the direction that a mechanical shift-share model would predict. But these effects are larger than the mechanical model would generate, indicating the presence of amplifying indirect effects of the age distribution on unemployment.

JEL Classifications: E24, J21
Keywords: age distribution, unemployment, demographics, shift-share

---

Bruce Fallick is a senior vice president in the research department at the Federal Reserve Bank of Cleveland. His email address is bruce.fallick@clev.frb.org. Christopher L. Foote is a senior economist and policy advisor in the research department of the Federal Reserve Bank of Boston. His email address is chris.foote@bos.frb.org.

The authors thank Joshua Ballance, Anne Chen, Emily Dohrman, and Luke Stewart for outstanding research assistance. They also received helpful comments from participants at the Federal Reserve System's 2022 Applied Micro Research Conference.

This version: October 2022 **https://doi.org/10.29412/res.wp.2022.15**

1 Introduction

Analysts have long recognized that shifts in the age distribution, most importantly changes in the share of young persons in the population, can affect the aggregate unemployment rate and its interpretation. Young persons tend to have higher unemployment rates than older people, so an increase in the youth share of the population should mechanically and directly raise the overall unemployment rate.

Since at least Perry (1970), the most common way to gauge the influence of the age distribution on the unemployment rate has been through the use of a "shift-share" method, in which age-specific unemployment rates are held constant at a base level while labor force or population shares change over time (or vice versa). More recent studies—Aaronson et al. (2015), Tuzemen (2017), and the Congressional Budget Office (2019)—use the shift-share method to argue that the aging of the labor force and other demographic factors have lowered the natural rate of unemployment in the United States since the turn of the 21st century, with important implications for US wage inflation and monetary policy.[1]

However, the shift-share method implicitly assumes that age-specific unemployment rates are unaffected by changes in the age distribution; that is, one group's population share has no indirect effects on the unemployment rate of another group. As Shimer (1998, 2001), among others, points out, this assumption may be problematic. Our paper uses state-level data to revisit the potential indirect effects of changes in the age distribution for the unemployment rate. We make four improvements upon previous analyses.

First, our regression specification takes the full age distribution into account. Most previous analyses concentrate on the youth share of the population or labor force, because this group has much higher-than-average unemployment rates and the aging of the baby boomers has caused the youth share to change significantly over time. However, over our sample period, there have been notable changes in other parts of the age distribution as well.[2] We therefore construct a "demographic index" that represents the mechanical influence of changes in the full distribution of age groups. As explained below, the coefficient on this index should indicate the presence and direction of any indirect effects arising from changes in the entire age distribution.[3] Following previous work on the economic effects of demographics, in some specifications we use information on lagged population structures to construct instrumental variables for a subset of our sample period. Doing so reduces the

[1] Shackleton (2018, Appendix B) outlines the method the Congressional Budget Office uses to demographically adjust the unemployment rate. Barnichon and Mesters (2018), Cajner and Cairo (2013), and Hetze and Ochsen (2006) present alternative methods for dealing with demographic change.

[2] Neumark and Yen (2020) also argue that changes in age-group shares aside from those of young persons may have indirect effects on unemployment.

[3] Fair and Dominguez (1991) and Skans (2005) take a different approach, including the population shares of every age group on the right-hand side of the regression.

possibility that our results are contaminated by the potential migration of certain population groups to low-unemployment states.

Second, we take advantage of nearly two decades of additional data on state-level age shares and unemployment since the original state-level analysis of Shimer (2001). Given the low amount of variation present in state-level demographic data once national trends are removed, these additional years are particularly valuable. We also try to add some years at the beginning of the sample (that is, the late 1960s and early 1970s) in order to pick up as much of the rise and fall of state-level youth shares as possible.

Third, consistent with recent approaches in labor market analysis, we estimate the state-level unemployment regressions as dynamic models, that is, with lagged unemployment rates on the right-hand side.[4]

Finally, we use econometric techniques designed for settings where spatial and serial correlation are significant problems. As Foote (2007) points out, there is ample evidence that state-level unemployment rates are cross-sectionally correlated, even after these rates are demeaned with respect to state and year averages. This correlation requires appropriate methods to construct standard errors; we use the method of Driscoll and Kraay (1998) throughout the paper. In some regressions, we also employ spatial-panel techniques that provide more efficient estimates when significant cross-sectional correlation is present.

Our analysis yields two main findings. First, the age distribution affects the unemployment rate in the direction that a shift-share analysis would predict; for example, a large cohort of young persons raises the unemployment rate. Second, there is evidence of indirect effects of the age structure that amplify its direct effects. That is, the indirect effects are in the same direction as the direct effects, so combining both direct and indirect effects generates a total demographic effect on unemployment that is larger than that of a shift-share model alone.

Our finding of demographic effects that are larger than those in the shift-share model has implications for both policymakers and researchers. On the policy front, several analysts recently have made a series of downward revisions to estimates of the US natural rate of unemployment. A potential explanation for these revisions is that demographic effects at the national level are pushing the national unemployment rate lower than the shift-share method would imply. For researchers, our results raise questions about the nature of the implied indirect effects and how to incorporate them into models of the labor market.

The paper proceeds as follows. Section 2 describes the standard shift-share analysis and some previous work that attempts to look beyond it. Section 3 introduces our regression specification, which includes a novel index that is intended to summarize the implications of a shift-share analysis throughout the age distribution. Section 4 describes the data. Section

[4]Shimer (2001) uses AR(1) corrections instead.

5 presents our main results, which use OLS, instrumental variables (IV), and spatial-panel methods. Section 6 further investigates the indirect effects by estimating the influence of the age distribution on unemployment rates within age groups. Section 7 recounts several hypotheses from the literature that may contribute to the indirect effects, and section 8 concludes.

2 The Age Distribution and the Unemployment Rate

2.1 Shift-Share Analysis

As noted above, the shift-share approach has been a standard method of accounting for the influence of changes in the age distribution (or other dimensions of demographics) on the unemployment rate. Such an analysis proceeds from the simple observation that the aggregate unemployment rate can be written as a weighted sum of the unemployment rates within age groups, with the weights being the labor force shares of the age groups. Thus, one can write

$$UR_t = \sum_a UR_{a,t} \cdot (LF_{a,t}/LF_t), \tag{1}$$

where the subscripts a and t refer to age group and year, respectively; UR_t is the overall unemployment rate in year t; and $UR_{a,t}$ is an age-group-specific unemployment rate in year t. The labor force levels LF_t and $LF_{a,t}$ are defined analogously. This decomposition shows that changes in the age distribution of the labor force mechanically imply changes in the aggregate unemployment rate. A simple shift-share analysis gauges this mechanical effect by holding age-specific unemployment rates constant and allowing labor force shares to vary over time.

Figures 1 and 2 illustrate the potential for shifts in the age distribution to affect the national unemployment rate in this fashion. Figure 1 depicts age-group-specific rates of unemployment since 1960. The unemployment rate regularly exceeds 15 percent for US 16- to 19-year-olds but is generally less than 5 percent for prime-age and older groups. Moreover, differences in unemployment rates across age groups have been fairly stable over the 60-year period. Figure 2 illustrates how the age distribution has shifted over time. The most obvious change, of course, has been the movement through the age distribution of the large baby boom cohort, whose members were born from 1946 through 1964. The shift-share method applies the changes in the age distribution from Figure 2 to the unemployment rate differentials in Figure 1.[5] If unemployment rates within age groups are unaffected by the

[5]Here, we abstract from differences in labor force participation rates, which we account for in our actual application below.

relative sizes of different groups, then the implications for the aggregate unemployment rate are straightforward.

2.2 State-Level Analysis and Indirect Effects

A shift-share analysis provides a simple way to analyze the direct effects of changes in the age distribution on the national unemployment rate. But time series analysis at the national level does not provide much scope for detecting indirect effects, which may include responses by employers, workers, and institutions to the changing demographic landscape, and which may work in the same direction as the direct effect or in the opposite direction.[6]

As in many areas of economic research, using state-level data can overcome some of the limitations of national data.[7] Shimer (2001) and Ochsen (2021) argue that the indirect effects of changes in the age distribution are significant and that such effects can be identified in a state-level study. Shimer (2001) concentrates on changes in the youth share.[8] That study uses panel data from US states to estimate an equation of the form

$$lnUR_{st} = \beta ln(Yshare_{st}) + \phi_s + \phi_t + \epsilon_{st}, \tag{2}$$

where UR_{st} is the unemployment rate in state s in year t, $Yshare_{st}$ is the share of the working-age population in state s in year t who are aged 16 to 24, and ϕ_s and ϕ_t are state and year fixed effects. The working-age population is defined as persons 16 to 64, and the sample period for most states is 1970 through 1996.[9] Both the unemployment rate and the youth share are entered as natural logs.

Shimer's estimate of β is negative. That is, a higher youth share reduces the aggregate unemployment rate, in contrast to the direction of direct mechanical effects. Shimer interprets this finding as evidence that an increase in the youth share of the population induces opposite-signed reactive changes in employer behavior (that is, indirect effects) that are strong enough to more than offset the direct effect of young persons' higher-than-average

[6] We review potential sources of indirect effects in section 7.

[7] One novel use of national data is a paper by Barnichon and Mesters (2018), who use a dynamic factor model to identify common variation in labor market flows across demographic groups at the national level. In this way, and in contrast to standard shift-share treatments, they allow both group-specific unemployment rates and the labor force shares that serve as weights in equation (1) to vary over time, as indicated by the demographic-specific components of the labor market flows. While the variation over time in demographic-specific unemployment rates may reflect indirect effects of demographic change, these are not explicitly addressed.

[8] Wachter (1977) and Korenman and Neumark (2000), among others, address the related question of the influence of cohort size on youth unemployment.

[9] The starting date for the states in Shimer's analysis is either 1970 or 1973, depending on the availability of unofficial BLS unemployment rates for the state. Official BLS rates are available from 1976 on. See the data appendix for details.

unemployment rates on overall unemployment.

Foote (2007) shows this negative coefficient to be fragile. Shimer's sample period ranges from the early 1970s through 1996; adding more years of data to the end of the estimation period causes the negative coefficient on the youth share to move toward zero. In the next section, we show how the youth-share coefficient evolves as data for the past two decades are added. We suggest some improvements to the regression equation that generate a more-stable estimate of demographic effects. And we develop a method of determining how changes in the entire age distribution—not just the youth share—affect unemployment.

3 Regression Specification

3.1 Dynamic Specification and Additional Cyclical Variables

The strong degree of serial correlation in state-level unemployment rates points to estimating a dynamic model, by placing a lagged dependent variable on the right-hand side. Alternatively, the researcher can use an AR(1) correction, such as the Prais-Winsten method, as in Shimer (2001). Dynamic models are more common in recent labor market research, and we argue below that they are appropriate in this context as well. If one continues to use the youth share as the demographic variable of interest, this gives

$$lnUR_{st} = \delta lnUR_{s,t-1} + \beta ln(Yshare_{st}) + \phi_s + \phi_t + \epsilon_{st}. \tag{3}$$

With a dynamic specification, the long-run effect of the youth share on unemployment is $\frac{\beta}{1-\delta}$. There is a large literature on the estimation of dynamic models with fixed effects that focuses on the downward bias in the coefficient on the lagged dependent variable (Nickell 1981). This bias, however, is on the order of $\frac{1}{T}$, where T denotes the number of observations available for each cross-sectional unit. Our regressions will typically include around 40 to 50 years per state, so we do not believe that this bias is a significant problem in our context.

Our regressions also include variables intended to soak up state-level demand shocks that could induce a reverse correlation between the demographic makeup of the population and the unemployment rate. Such correlations could arise, for example, if prime-age persons (who have low unemployment rates) tend to move to states that are booming. One set of these variables are well-known Bartik variables. These regressors are state-level weighted averages of national industry-level employment growth, with the industries defined by the Bureau of Economic Analysis for the national accounts. The weights for each Bartik variable in state s in year t are industry shares of gross state product in year t–1.[10]

[10] Additionally, the national industry-level employment growth rates used to construct the Bartiks are calculated relative to the national growth rate for aggregate employment. See the data appendix for details.

The other set of cyclical regressors exploits the propensity for housing construction to rise in states with booming economies. Our housing regressor is constructed by first dividing the state's total housing permits by the state's population, and then calculating the change in the log of this ratio from years t–1 to t.[11] Like the Bartiks, the housing variables proxy for labor demand, but they also sharpen the interpretation of the demographic coefficient. As young persons move out of their parents' houses and into homes of their own, overall housing demand rises (Mankiw and Weil 1989). Consequently, a surge in a state's youth share could induce a surge of housing demand that reduces unemployment due to an increase in construction (and any associated multiplier effects). By placing a construction variable directly in the regression, we hope to purge the coefficient on the demographic index of this potential confounder. For both the Bartik and housing variables, we include the contemporaneous value and two lags.

Figure 3 illustrates how the youth-share estimate changes with the addition of the past 20 years of data, the inclusion of the cyclical regressors, and the addition of a lagged dependent variable. The top-left panel of the figure shows the changes in the coefficient on the youth share estimated with a static model and a Prais-Winsten AR(1) correction. Ending the sample in 1996 generates a negative coefficient. As we extend the sample forward in time, the coefficient moves toward zero, although it remains negative. The top-right panel adds the cyclical regressors, and the point estimates of the coefficient rise to small positive values. The lower panel includes both the cyclical regressors and a lagged dependent variable, so the model is now dynamic. The estimated youth-share effect rises to positive values, and it becomes more robust to the extension of the sample period. We take the latter robustness as evidence that a dynamic model is preferable to an AR(1) correction, and we use dynamic models in our estimating equations below.

3.2 Demographic Index

Although the unemployment rate of young persons stands out in Figure 1 as much higher than those of other age groups, the differences among other age groups are not trivial. We therefore want to account for movements in population shares throughout the entire age distribution. To do so, we construct a summary statistic that we call a *demographic index*, which embodies the full implications of a shift-share analysis. This index will replace the youth share in our estimating equations.

Recall from equation (1) that the unemployment rate may be decomposed as a weighted average of age-specific unemployment rates, where the weights are labor force shares. The

[11]See the data appendix for details.

same, of course, is true at the state level:

$$UR_{s,t} = \sum_a UR_{a,s,t} \cdot (LF_{a,s,t}/LF_{s,t}), \tag{4}$$

where the subscript s refers to the state. To relate demographic shifts to shares of *population* rather than shares of the *labor force*, we introduce labor force participation rates into the analysis. Define

$$v_{a,s,t} = UR_{a,s,t} \cdot (LFPR_{a,s,t}/LFPR_{s,t}), \tag{5}$$

where $LFPR$ denotes either an overall or age-specific participation rate as indicated. Then we have

$$UR_{s,t} = \sum_a popshare_{a,s,t} \cdot v_{a,s,t}, \tag{6}$$

where $popshare$ denotes an age group's population share.

Using this formulation, we can construct two indexes that capture the mechanical shift-share effect of the age distribution on the aggregate unemployment rate. The first is a *fixed-weight* index, defined as

$$fw_{st} = \sum_a popshare_{a,s,t} \cdot \overline{v_{a,s}}, \tag{7}$$

where $\overline{v_{a,s}}$ is the mean of $v_{a,s,t}$ for each age group in each state over the entire sample period.

While straightforward, the fixed-weight index does not vary as relative labor force participation or unemployment rates change over time. Most notably, for much of our sample period, the participation rates of young persons trended down, and those of women trended up. To explicitly account for such changes, we also calculate a *chain-weight* index.[12] Define

$$dcw_{s,t} = \sum_a (popshare_{a,s,t} - popshare_{a,s,t-1}) \cdot \left(\frac{v_{a,s,t} + v_{a,s,t-1}}{2}\right). \tag{8}$$

The chain-weight index is then

$$cw_{st} = \sum_0^t dcw_{s,t} + \overline{ur_{s,t}}, \tag{9}$$

where the mean aggregate unemployment rate over the sample period, $\overline{ur_{s,t}}$, normalizes the index for easier comparison to the actual unemployment rate.

If there are no indirect effects of changes in the age distribution on the unemployment rate, then in principle the full influence of changes in the age distribution should be encompassed

[12] See Shimer (1998) for an earlier use of a chain-weight approach.

by these indexes.

With these indexes replacing the youth share in the unemployment equation, as well as the addition of the Bartik and housing variables, our main regression specification becomes

$$UR_{st} = \delta UR_{s,t-1} + \gamma Index_{st} + \Sigma_{k=0}^{2}\theta_k^B Bartik_{s,t-k} + \Sigma_{k=0}^{2}\theta_k^H Housing_{s,t-k} + \phi_s + \phi_t + \epsilon_{st}, \quad (10)$$

where $Index_{st}$ is either the fixed-weight index (fw_{st}) or the chain-weight index (cw_{st}). Note that neither the unemployment rate nor the demographic index is in logs. Consequently, in the absence of any indirect effects of the age distribution on unemployment, the long-run effect of the index on the unemployment rate, measured as $\frac{\gamma}{1-\delta}$, should equal one. A long-run coefficient greater than one would indicate indirect effects that tend to amplify the direct effects. Conversely, a coefficient less than one would indicate indirect effects that subtract from direct effects.

4 Data

4.1 Sources

We confine our sample to the 48 contiguous US states. Calculation of state-level fixed-weight and chain-weight indexes requires data on population, unemployment rates, and labor force participation rates by age, state, and year. State-level unemployment rates and the data necessary to construct the demographic indexes from 1978 onward are easily obtainable. However, as is clear in Figure 2, starting the sample in 1978 misses an earlier period when the population shares of the younger age groups were rising. By extending the indexes back to 1963, we provide potentially valuable additional identifying variation, though at the cost of some diminution in data quality. Recognizing the trade-off, we will also run regressions that use samples beginning in 1963 with samples that begin in 1978.

The data appendix outlines the construction of our data set in detail; here are the important takeaways:

- **State-level unemployment rates.** Official state unemployment rates from 1976 on are available from the Bureau of Labor Statistics' Local Area Unemployment Statistics (LAUS) Program.[13] The BLS has produced unofficial state-level estimates of annual unemployment rates back to 1970 for 38 larger states and back to 1973 for smaller states; the BLS supplied these estimates to Shimer for his 2001 paper, and he kindly passed them on to us. Our state-level unemployment rates from 1963 to either 1970

[13] See https://www.bls.gov/lau/.

or 1973 are based on rates constructed by state agencies, largely from unemployment-insurance (UI) data, and published in the US Labor Department's *Manpower Report of the President* (later named the *Employment and Training Report of the President*). When these rates were calculated, they were standardized to some extent by the BLS, which provided a method by which UI-based data could generate unemployment rates that were as comparable as possible to unemployment rates from the Current Population Survey (CPS). As described in the data appendix, we spliced these early rates to the unofficial BLS estimates from the early 1970s to obtain state-level unemployment rates back to 1963.

- **Age- and state-specific unemployment and labor force participation rates.** These rates are required for constructing the demographic indexes, and we also place age-/state-specific unemployment rates on the left-hand side of some regressions. We obtain data on unemployment rates and labor force by age, state, and year from the LAUS.[14] These data are not published by single year of age but rather for the following age groups: 16 to 19, 20 to 24, 25 to 34, 35 to 44, 45 to 54, 55 to 64, and 65-plus. Accordingly, we use these age groups rather than single years of age to construct the demographic indexes. LAUS supplies these data from 1978 onward, so it is straightforward to calculate our demographic indexes starting in 1978. To extend the demographic indexes back to 1963, we impute age-specific unemployment and labor force participation rates for those years by regressing unemployment and participation rates for each state, age group, and gender on corresponding national rates for the 1978–2019 sample. We then apply the estimated coefficients to the national rates for each age group and gender for the 1963–1977 period.

- **Age-specific population data.** We take data on population by age, state, and year from the National Cancer Institute's Surveillance, Epidemiology and End Results (SEER) Program (SEER 2022). These data start in 1969.[15] SEER data for state-age groups are generally preferred to data published by the US Census Bureau because the SEER data are smoothed through the occasionally abrupt changes that can appear in census data in decennial years. For years before 1969, which are not covered by the SEER, we construct interpolated estimates of state-level population using decennial census data. As described in the data appendix, we use the census data to estimate age-specific populations for 1960 and 1965; the latter year's estimates come from applying 1965–1970 migration rates to microdata from the 1970 decennial census. Combining these estimates with the SEER data from 1969 gives us three years (1960, 1965, and

[14]Inspection of these data reveals that they are very similar to raw age-group means calculated from CPS microdata.

[15]See https://seer.cancer.gov/popdata/.

1969) between which to linearly interpolate age-specific yearly population data by state during the 1960s.

4.2 Basic Patterns in Demographic Indexes

We now provide a sense of what the demographic indexes would look like at the national level and how they relate to the US unemployment rate. Figure 4 depicts a fixed-weight demographic index constructed with quarterly data for the nation as a whole to the actual US unemployment rate and to the non-cyclical rate of unemployment from the Congressional Budget Office (sometimes known as the CBO NAIRU, for *non-accelerating inflation rate of unemployment*). The demographic index (solid blue line) rises and falls during the past 60 years broadly in line with the trend in the actual unemployment rate (red dashed line). However, the CBO's estimate of the trend unemployment rate (green dashed line) is amplified relative to what the changes in the age distribution would mechanically imply.

The quarterly demographic index in Figure 4 is calculated with national labor force, unemployment, and population data from the BLS. Figure 5 depicts annual fixed-weight and chain-weight indexes constructed with SEER population data and our imputed population data, the raw materials for our state-level demographic indexes. The vertical gray line denotes 1969, the year that the SEER population data begin. The figure shows that our interpolations of population before 1969 generate a somewhat choppier series. However, a comparison of this figure with the previous one shows that the basic patterns in demographic indexes remain similar regardless of the source of population data. Note also that the chain-weight index (red dashed line) declines by more than the fixed-weight index (blue line) from the early 1980s onward. Because the chain-weight index allows age-specific labor force participation rates to change, the greater decline in the chain-weight index reflects the falling participation rates of (high-unemployment) young workers in the US labor force and the rising participation rates of (low-unemployment) prime workers, specifically prime-age women.[16]

Both of the indexes plotted in Figure 5 reflect the rise and fall of the youth share of the US population, but at the state level, the relationship between the youth shares and the demographic indexes is less straightforward, especially when state and year means have been removed from both series. Figure 6 plots fixed-weight indexes against youth shares for six selected states. Both the fixed-weight indexes and the youth shares have been demeaned

[16] We found that state-level versions of fixed-weight and chain-weight indexes often vary by a few tenths of a percentage point. These differences are generated by differences in mean age-specific unemployment and participation rates relative to corresponding rates at the beginning of the sample (the "jumping off" point for the chain-weight index). The mean adjustment added to the chain-weight index was not calculated to equate the fixed-weight and chain-weight indexes, so these differences resulted in small differences in the final indexes.

from state and year averages. Trends in the two series are broadly similar in each panel, but there are periods when the youth share fluctuates by much more than the demographic index does. This additional variation likely arises because in Figure 6, the youth share is defined as in most previous research, that is, as the share of 16- to 24-year-olds relative to the *working-age* population, not *total* population. Because the working-age population is defined as persons aged 16 to 64, the aging of a large group of persons from their early 60s to their late 60s will impart variation to the youth share by significantly reducing its denominator. Such aging imparts less variation to the demographic indexes; the older age groups are included in these indexes, although they are downweighted in line with their lower labor force participation rates.

4.3 Initial Evidence on Demographics and Unemployment

How do the demographic indexes correlate with unemployment at the state level? Figure 7 provides some initial answers by plotting demeaned fixed-weight indexes against demeaned unemployment rates for the same six states that were featured in the previous figure. For some states, such as West Virginia and Louisiana, the fixed-weight index and the unemployment rate trend together over time. In other states, such as South Carolina and North Dakota, a positive correlation is less apparent.

Figure 8 brings all states into the analysis by comparing long differences of fixed-weight indexes and unemployment rates at the state level. The upper-left panel shows the changes from 1963 to 2019, the endpoints of our longest possible sample period. There is a positive relationship between index changes and the unemployment-rate changes, as we would expect, although a small state (Vermont) appears as an outlier. The remaining panels depict long differences during various sub-periods, with the top-right panel depicting changes from 1963 to 1979. Here, a positive relationship is much less apparent, although two small states (Vermont and Wyoming) appear as influential data points that pull the relationship in a negative direction.

The two panels in the bottom row of Figure 8 depict data for the 1979–1999 and 1999–2019 sub-periods. These panels are particularly informative because the BLS provides official state-level unemployment rates from 1976 onward. Consequently, the changes in unemployment rates in these two panels are a true apples-to-apples comparison of similarly constructed rates in two different years. This similarity stands in contrast to the panels in the top row, which use UI-based unemployment rates from 1963 as the beginning years for their comparisons and official BLS rates for the ending years. In any case, the panels in the bottom row suggest a positive relationship between index changes and unemployment-rate changes, as we would expect.

Table 1 analyzes the long differences depicted in the four panels more formally using regressions. Odd-numbered columns are unweighted, while even-numbered columns are weighted by the natural log of the state's mean population from 1963 through 2019. In the first two columns, which use the entire 1963–2019 sample period, the coefficient on the index is a positive and significant 1.6 in both the weighted and unweighted regressions. The next two columns, which use the 1963–1979 sub-period, show negative coefficients, and weighting by population reduces the coefficient from –2.0 to –1.3 while also substantially reducing the standard error. According to Solon, Haider, and Wooldridge (2015), large changes in coefficients and standard errors when using weights are tell-tale signs of heteroskedasticity. In this case, the problem probably arises because the unemployment rates for small states are likely to be less accurate before 1976, the year that official BLS unemployment rates became available. If so, then the extra variance of the small-state residuals reflects increased measurement error.[17] The difficulty in obtaining accurate readings on small-state unemployment is likely to be especially acute for the years before 1969, when we must infer unemployment rates from the UI-based rates published in the *Manpower Report of the President*. We view these 1963–1979 results as reasons to exercise caution when interpreting results that are based on pre-1976 data. In any case, much of the analysis below uses data only from years for which official unemployment rates from the BLS have been published.

The remaining columns of Table 1 use BLS unemployment data throughout. For both the 1979–1999 and 1999-2019 sub-periods, the estimated coefficients on the fixed-weight index change are positive, and neither the coefficients nor the standard errors are materially affected by population weighting. Notably, all of the positive coefficients in Table 1 are equal to one or greater, although in no case is the regression precise enough to reject the hypothesis that the coefficient equals the value implied by the shift-share model. In the next section, we present regressions that make much better use of the available data. In many cases, the regressions do support the idea that the effect of demographics on unemployment is larger than the shift-share model would predict.

5 Regression Results

5.1 OLS and IV

Table 2 presents OLS estimates from the dynamic panel model specified in equation 10, with and without the Bartik and housing cyclical controls.[18] Columns 1 through 4 present

[17]Recall that Vermont and Wyoming appear as influential observations in the 1963–1979 panel of Figure 8. Removing those two states from the regression for this sub-period (not shown) causes the 1963–1979 index coefficient to rise to a positive 0.57 (s.e. = 1.55) in an unweighted regression and to a positive 0.70 (s.e. = 1.49) in a population-weighted regression.

[18]As noted earlier, to account for spatial and serial correlation, we use Driscoll-Kraay standard errors. The bandwidth for these errors is five years. In section 5.2, we use techniques from the spatial-economics

the estimates using the fixed-weight demographic index, while columns 5 through 8 use the chain-weight index. Each regression in the table uses the full 1963–2019 sample.

Column 1 includes no cyclical controls and reveals a significant coefficient of 0.427 on the fixed-weight demographic index. Given a point estimate of 0.845 on the lagged dependent variable, this estimate implies a long-run effect of the demographic index (in bold) of 2.76, much larger than the value of one implied by the shift-share hypothesis. The standard error on the long-run effect is large, however, so a value of unity cannot be rejected. The three Bartik variables are added in column 2, and the three housing variables are added in column 3. These additional regressors have modest effects on the point estimates of the demographic index, but this estimate is significant only at the 10 percent level when all six controls are added. The long-run effect of demographics in column 3 remains above 2, but the standard error associated with this estimate remains large enough to prevent us from rejecting a long-run effect of unity. Column 4 weights the regression with the log of mean population from 1963 through 2019, which produces only minor changes. Columns 5 through 8 repeat the analysis using the chain-weight index rather than the fixed-weight index. The results are similar, with point estimates of the long-run effect much greater than one, yet the standard errors remain large enough to prevent us from rejecting a long-run effect equal to one. All told, the results in Table 2 are suggestive of indirect effects that amplify the direct effects, but we cannot reject the hypothesis that there are no indirect effects.

A well-known issue with estimating demographic effects at the state level arises from endogenous migration. If particular age groups are more likely to move to states with low unemployment rates or high rates of job growth, then the resulting change in the state's demographic makeup could distort estimates of the causal effect of demographics on state-level unemployment. Our Bartik and housing variables are designed to soak up this type of variation, but they may not account for all of it. Previous research accounts for endogenous migration by using lagged birth rates or past demographic patterns to instrument for current demographic structures. This approach makes use of the fact, for example, that the number of 35-year-olds in a state today is closely related to the number of 25-year-olds living there 10 years earlier.[19] As described in the data appendix, we use the age distribution of the state lagged by 10 years to construct an index that serves as an instrument for the current demographic index. We construct this lagged-demographics instrument using our detailed population data from SEER that begin in 1969. Consequently, we can estimate the IV model only for a sample that begins in 1979.

---

literature that are more efficient when spatial correlation is present.

[19] See Shimer (2001), Karahan, Pugsley, and Şahin (2021), Maestas, Mullen, and Powell (forthcoming) and Crump et al. (2019) for uses of lagged demographic information to instrument for current demographic structure.

Our IV regressions are reported in Table 3.[20] Column 1 presents the OLS results on the shorter sample using the fixed-weight index. The estimated long-run effect from the 1979–2019 sample is in excess of 3, much larger than the estimated long-run effect using OLS on the longer sample in Table 2. The key point, however, is that when the same regression is estimated with IV, as in column 2, the resulting estimate of the long-run effect (3.16) is quite close to the OLS estimate (3.43). Moreover, like the OLS estimates in the previous table, the IV estimate changes little with population weighting in column 3. This similarity between OLS and IV estimates gives us confidence that endogenous migration is not biasing the results in this shorter sample and therefore probably not in the full sample in Table 2 either.

Columns 4 through 6 of Table 3 present analogous results using the chain-weight index. The point estimates for the long-run effects are even larger than those using the fixed-weight index, but they are not significantly affected by the use of IV or by population weighting. In the interest of brevity, we confine ourselves to the fixed-weight index in the remainder of the paper.

5.2 Spatial Models

A basic challenge with any state-level study of demographics is that relatively little demographic variation remains after national trends have been absorbed by year fixed effects and cross-state-level differences have been absorbed by state fixed effects. Notably, although the baby boom generated a great deal of national variation in age shares over time, the amount of regional variation it brought about was much smaller, because the increase in births closely followed World War II in all states. Making matters worse, the demographic variation that does remain after accounting for national trends includes large amounts of both spatial and serial correlation. Demographic shifts occur slowly within states over time (serial correlation), while neighboring states experience similar demographic shifts at the same time (spatial correlation). Both serial and spatial correlation are also found to lesser degrees in state unemployment rates. Among other issues, the simultaneous presence of spatial and serial correlation complicates the calculation of standard errors in a study of demographic effects on unemployment (Foote 2007).

So far, we have dealt with spatial correlation in unemployment rates and demographic patterns by relying on robust standard errors calculated with the method of Driscoll and Kraay (1998). In this section, we estimate spatial-panel models that can use the information in the state-level data more efficiently. Spatial models require the researcher to specify an exogenous $N \times N$ weighting matrix, where N is the number of cross-sectional units (here,

[20]For brevity, this table and the remaining ones omit the estimated coefficients on the Bartik and housing controls, although they are included in the regressions.

the 48 contiguous US states). The matrix should reflect the form of spatial correlation in the residuals. As is common in state-level analysis, correlations in both demographics and unemployment between states that border one another tend to be stronger than correlations between states that do not. Accordingly, our weighting matrix is based on a first-order contiguity matrix, where the (i,j)'th entry of the weighting matrix equals one if state i and state j share a common border and zero otherwise.[21]

We use this matrix in three ways. First, we assume that the residuals in our regression are correlated as the first-order continuity matrix would suggest. The regression can then use this information to more efficiently estimate the coefficients, just as generalized least squares (GLS) is an improvement over OLS. Second, we assume that a state's unemployment rate depends in part on a weighted average of nearby states' unemployment rates. Because the spatial weighting matrix supplies these weights, we are assuming that a state's unemployment rate can be affected by the average unemployment rate of the states that border it. These first two uses of the weighting matrix imply that we are estimating both a combined spatial error model (SEM) and a spatial autoregressive model (SAR). Third, we also use the weighting matrix to construct weighted averages of other states' fixed-weight demographic indexes. This so-called spatial lag of the fixed-weight index allows us to determine whether a state's unemployment rate depends not only on its own demographic index, but also on the weighted average of bordering states' indexes.[22] The final spatial model is

$$\begin{aligned} UR_{st} = {} & \delta UR_{s,t-1} + \rho(\Sigma_{j=1}^{N} w_{ij} UR_{jt}) + \\ & + \gamma Index_{st} + \mu(\Sigma_{j=1}^{N} w_{ij} Index_{jt}) \\ & + \Sigma_{k=0}^{2} \theta_k^B Bartik_{s,t-k} + \Sigma_{k=0}^{2} \theta_k^H Housing_{s,t-k} + \phi_s + \phi_t + \epsilon_{st}, \\ \epsilon_{st} = {} & \lambda(\Sigma_{j=1}^{N} w_{ij} \epsilon_{jt}) + \nu_{st}, \end{aligned}$$

where w_{ij} is the element in row i and column j of the exogenous $N \times N$ weighting matrix W. In this framework, ρ is the spatial autoregressive coefficient, which picks up the impact of bordering-state unemployment on home-state unemployment. The spatial error coefficient λ measures the degree of correlation in the errors of states that share a common border. The coefficient on the spatial lag of the demographic index μ measures any correlation between bordering-state demographics and home-state unemployment that remains after accounting for home-state demographics directly.

The results of our spatial model using the fixed-weight index appear in Table 4. We

[21]As is common in the spatial literature, this matrix is row-normalized so that the sum of the elements in each row equals one. This normalization causes spatial effects to reflect weighted averages of variables in surrounding states.

[22]Alternative weighting matrixes based on rates of bilateral migration yield similar results, as do matrixes based on similarities in industry- and occupational-level employment shares across states.

use the estimation method of Lee and Yu (2010), which has been coded into Stata as the program `-xsmle-` by Belotti, Hughes, and Mortari (2013).[23] The first column of Table 4 reproduces our basic OLS results for the full sample period from column 3 of Table 2. The second column estimates the combined SAR and SEM model. The third column adds the weighted average of the fixed-weight index from neighboring states to the list of regressors. The SEM and SAR terms enter significantly, while the spatially lagged demographic index does not.[24] The estimated long-run effects from the spatial models are, like those from the OLS models, well above 2.

The value of the spatial model can be seen by comparing the estimated demographic coefficient using OLS in column 1 to the coefficient estimated by the spatial model in columns 2 or 3. The estimate is 0.399 using OLS, which is significant at the 10 percent level. With the spatial model of column 2, however, the point estimate rises to 0.504, which is significant at the 1 percent level. Because both the OLS and spatial models use Driscoll-Kraay standard errors with the same lag length (five years), comparing the significance levels in the two models is an apples-to-apples comparison. The modestly smaller standard error in the spatial model reflects the gain in efficiency that is achieved when the estimation procedure takes spatial effects into account.

The spatial model also provides evidence that the demographic index picks up true demographic effects rather than a confounding factor that is correlated across states. Consider the estimated coefficient on the spatial lag of neighboring states' demographic indexes, which equals an insignificant –0.316 in column 3. As noted, demographic factors and potential confounders are likely to be correlated across states. If the home-state demographic coefficient is large and positive because of a confounding factor, then this shared spatial correlation should cause the neighboring-state demographic coefficient to be large and positive as well. The small negative value of the neighboring-state coefficient indicates that once we condition on neighboring-state unemployment rates via the spatial autoregressive term, the demographic effects stop at the border. This finding increases our confidence that demographics are exerting a true causal effect on unemployment.

Column 4 of Table 4 reproduces the basic OLS results for the shorter sample period of 1979 through 2019 from column 1 of Table 3. Column 5 shows the coefficients from a model without the spatial lag of bordering-state indexes, while column 6 includes this variable. As in columns 1 through 3 of the table, the spatial models increase the statistical significance of

[23]See Belotti, Hughes, and Mortari (2017) for a description of this procedure.

[24]Specifically, the spatial autoregressive coefficient in column 2 is a statistically significant 0.114. This point estimate indicates that home-state unemployment rises by 0.114 percentage points when the weighted average of unemployment in surrounding states rises by 1 percentage point. The corresponding spatial error coefficient is 0.218, which is also statistically significant. This estimate reflects the degree of correlation in neighboring-state error terms. Like GLS, the regression uses this correlation to improve the efficiency of the estimates.

the demographic-index coefficients. They also yield estimates of the long-run effects similar to those estimated by OLS, although the estimates from the spatial regressions are more precise.

The bottom panel of Table 4 includes p-values for the hypothesis that the long-run effect of the demographic index equals one, as would be implied by a shift-share analysis. In all of the spatial models, we can reject this hypothesis at the 10 percent level, and we can reject it at the 5 percent level in three of the four spatial models. This strengthens the indications from the earlier OLS and IV results that there are indirect effects of the age distribution that amplify the direct effects, which in turn implies that the age distribution affects age-specific unemployment rates. In the next section, we test this hypothesis directly by examining the relationship between our demographic index and the unemployment rates of individual age groups.

6 Age-Specific Results

6.1 Regressions Using Individual Age Groups

The finding that changes in the age distribution affect the unemployment rate by more than the shift-share model would predict implies that the age distribution affects the unemployment rates of individual age groups. A straightforward way to check this prediction is to regress the unemployment rates for specific age groups on the demographic index. Positive coefficients would imply the presence of indirect effects of the age distribution that amplify the direct effects of the shift-share model.

Age-group-specific unemployment rates by state can be calculated from CPS microdata starting in 1978.[25] Given the dynamic specification of our regression, this limitation translates into regression samples that start with 1979 due to the inclusion of a lagged dependent variable in the model. Because sample sizes become smaller as one disaggregates by age as well as by state, sampling error is a larger concern with age-specific regressions than with statewide regressions. The unemployment rates calculated for the 10-year age groups that we used in producing our demographic index appear uncomfortably noisy, so we instead present regressions for three broader age groups that are commonly used in labor market research: young workers (aged 16 to 24), prime-age workers (aged 25 to 54), and older workers (aged 55 and older).[26] The mean level of the unemployment rate differs substantially between these age groups, so we estimate the age-specific regressions using the natural logs of both the unemployment rates and the demographic index. Doing so allows us to compare the

[25] As noted above, rates for 10-year age groups from 1978 onward are also published by the BLS, but we found that these rates are essentially identical to the rates we generated ourselves with CPS microdata.

[26] The demographic index is, as before, calculated using 10-year age groups.

proportionate impacts of the demographic index on each of the age groups more easily.

Table 5 presents our age-specific unemployment regressions. The odd-numbered columns of Panel A present the OLS results, which generate index coefficients that range from 0.181 for young workers to 0.414 for prime-age workers. None of the OLS estimates is statistically significant, however. The even-numbered columns in Panel A present results from spatial-panel models, and the point estimates are generally larger than those estimated for the same age group with OLS. The precision of the estimates is also improved, and the coefficient for the prime-age group becomes significant at 0.521.

Panel B shows IV regressions that use the same instrumenting strategy for the fixed-weight index as in Table 3. A notable finding here is that the coefficient for young workers increases from 0.181 under OLS to a highly significant 0.809 with IV. The prime-age coefficient also increases, from 0.414 to 0.484, but the latter coefficient is significant only at the 10 percent level. The coefficient for older workers also increases, from 0.353 to 0.749, but the large standard error for this age group prevents the coefficient from being statistically significant.[27] This pattern of positive coefficients is generally consistent with our earlier finding that at the aggregate level, increases in the demographic index have a larger effect on the unemployment rate than a shift-share model would predict.[28] Yet, the low statistical significance of many of the estimated index coefficients calls for a more efficient way of using the available data.

6.2 Statewide Averages of Age-Specific Effects

In this section, we pursue an alternative method of testing for demographic effects that is likely to be more efficient than the age-specific regressions. The basic idea is to partial out the *direct* effect of demographics from state-level unemployment rates and then ask whether the demographic index affects state-level unemployment once these direct effects have been removed.[29]

Consider an individual-level regression of the probability of unemployment ($Pr(U_{pst})$) for

[27]Note that for older workers, the standard error on the demographic-index coefficient is lower under IV than under OLS (0.559 versus 0.576). This decline probably reflects the very high relevance of the instrument (note the first-stage F-stats in Table 3) combined with a large amount of sampling error in unemployment rates for older workers. In private correspondence, a BLS official told us that sampling error is likely to be much larger for older workers due to their low participation rates and correspondingly small numbers of people in the labor force.

[28]Analogously, Crump et al. (2019) find that the age distribution affects unemployment-inflow rates within age groups, Engbom et al. (2019) find that the age distribution affects employer-to-employer rates within age groups, and Karahan and Rhee (2017) find that the age distribution affects migration rates within age groups.

[29]The method we will use is similar to the "demographic adjustment" for unemployment or wages that is often used in labor market analysis.

person p in state s in year t, where we condition on being in the labor force:

$$Pr(U_{pst}) = logit(\beta_t X_{pst} + \psi_{st} + \epsilon_{pst}), \tag{11}$$

where X is a vector of demographic variables, including age group, gender, education, and race.[30] The ψ_{st} term is an effect that is common across all persons in the same state and year, and ϵ_{pst} is a person-level residual. This equation is easily estimated with yearly regressions using CPS microdata from 1978 onward.

The β_ts are intended to capture the direct effect of demographics, most importantly the direct effect of age, which is fully represented by the coefficient on one's own age. The key to the greater efficiency of this model, at the cost of some realism, is to assume that the coefficients that describe these direct effects do not vary by state.[31] Estimating the model by year allows the β_ts to change each year but forces these estimates to be the same for each state in a given year. For a person of a given age, the influence of their state's age distribution on their probability of unemployment—that is, the indirect effects of their state's age distribution—is relegated to ψ_{st}, which is an effect shared by all persons in the same state and year.[32] In addition to capturing any indirect effects of demographics on unemployment, the ψ_{st} term also captures two other influences on state-level unemployment: idiosyncratic state-level shocks and cyclical shocks that are correlated across states and engender cross-state correlation in unemployment.

This model can be used to construct state-level unemployment rates in which the direct effects of demographics have been partialled out. To see this, consider an unemployment rate that combines the estimated coefficients from the yearly regressions with the same set of X variables for all states and years:

$$\widehat{UR_{st}} = logit(\hat{\beta}_t \overline{X} + \psi_{st}), \tag{12}$$

where $\overline{X}$ reflects the average demographic structure across all states and years. Because $\overline{X}$ is constant, $\widehat{UR_{st}}$ varies across states within a year only because of ψ_{st}, that is, because of indirect effects of demographics, idiosyncratic shocks that are unique to the state, and cyclical shocks that are correlated across states.

[30]Specifically, the regression includes dummy variables for 13 age groups, two genders, and four education groups (less than high school, high school graduate, some college, and college graduate). These variables are interacted so that age-unemployment profiles can vary by gender and education. In addition, there are dummy variables for race (white, Black, Hispanic, and other).

[31]Note that we do not assume that age effects are constant across age groups when constructing our demographic index.

[32]If the βs were permitted to vary by state, then they would capture indirect effects as well, just as the age distribution in the state affects the age-specific unemployment rates in section 6.

Now consider a regression of $\widehat{UR_{st}}$ on the state-level demographic index:

$$\widehat{UR_{st}} = \delta\widehat{UR_{s,t-1}} + \gamma Index_{st} + \phi_s + \phi_t + e_{it}, \tag{13}$$

where the Bartik and housing variables are not shown for brevity. This regression is designed to test for indirect demographic effects and is identified using the same type of variation as the age-specific regressions above. Those regressions removed the direct effects of demographics by focusing on specific age groups and determining whether changes in the demographic index affected age-specific unemployment rates, in contrast to the shift-share hypothesis. Similarly, construction of $\widehat{UR_{st}}$ removes the direct effects of demographics by partialling them out with a first-stage logit regression, leaving the indirect effects to be folded into the ψ_{st} term.[33] In both types of regressions, the coefficient on the index should be zero if indirect demographic effects are absent *and* the index is uncorrelated with regional business cycles. As with the age-specific regressions above, we attempt to control for regional cycles so that the estimated index coefficients capture only the indirect demographic effects.

Table 6 shows estimates from equation 13. The left-hand-side variable in all regressions is $\widehat{UR_{st}}$, which we can call a "demographics-constant" unemployment rate. Column 1 presents an OLS regression in which the only method of controlling for regional business cycles is the inclusion of the Bartik and housing variables. The coefficient on the demographic index is positive (0.279) but not significant. Column 2 uses a spatial model similar to those used in section 5.2; by accounting for correlations with neighboring states, this model arguably goes further to control for regional business cycles. As in previous tables, the use of a spatial model increases the size of the estimated demographic-index coefficient and improves its precision. This coefficient rises to 0.397 and becomes statistically significant. Columns 3 and 4 use the predicted fixed-weight index based on past demographics—used as an IV in earlier regressions—as a right-hand-side variable in a reduced-form model. Estimates are similar to the previous two columns, with the index coefficient entering significantly in the spatial regression of column 4.

The last four columns of Table 6 take into account the fact that the demographics-constant unemployment rate is constructed with CPS microdata and is therefore likely to have more sampling error than the state unemployment rates used previously in this paper. Those rates (from the early 1970s onward) came from the BLS's LAUS program, which applies modeling and smoothing techniques that improve upon the unemployment rates that result by simply taking means of unemployment data directly from CPS microdata. Consequently, any unemployment rates calculated directly from that microdata, like our demographics-constant rates, are likely to have more measurement error than the LAUS rates. In the

[33] Note that the common variation arising from yearly changes in the estimated β_ts will be absorbed by the yearly dummies in equation 13.

last four columns of the table, we enter the difference between the LAUS rate and a raw CPS mean rate as a right-hand-side variable.[34] This difference is highly significant in all regressions and enters positively, as we would expect.[35] The inclusion of this difference also tends to strengthen the evidence of indirect effects of demographics, as coefficients on the actual and predicted fixed-weight coefficients increase relative to those in previous columns.

7 Nature of the Indirect Effects

Taking stock of our results so far, we have found:

1. A state-level unemployment regression that includes a demographic index for unemployment reveals that the effects of demographics on unemployment likely extend beyond the direct mechanical effects implied by the shift-share model.

2. The strength of the indirect effects appears to come about because the index changes unemployment rates within age groups in ways that amplify the direct effects. These types of age-specific effects stand in contrast to the shift-share assumption that age-specific unemployment is unaffected by shifts in population shares.

What accounts for this amplification? Several possibilities can be drawn from the literature. For the most part, the studies cited below do not address the implications of their findings for the unemployment rate, but such implications are plausible. The studies fall into roughly three categories.

First, the indirect effects may reflect how the age distribution affects the recruitment, production, and investment strategies of firms. These strategies may involve the effect of the age distribution on worker turnover (Shimer 2001; Engbom et al. 2019; Clark and Ritter 2020), entrepreneurial propensities (Karahan, Pugsley, and Şahin 2021; Engbom et al. 2019; Kopecky 2019; Liang, Wang, and Lazear 2018), the quality of worker-firm matches and the rate of employer-to-employer movement (Engbom et al. 2019), labor force participation rates (Acemoglu and Restrepo 2017; Eggertsson, Lancastre, and Summers 2019; Cosic and Steuerle 2021), or bargaining power (Glover and Short 2020). Reactions in fiscal policy are also possible (Sheiner 2014).

Second, the indirect effects may reflect the sluggish adjustment of employment to changes in productivity caused by the age distribution (Maestas, Mullen, and Powell forthcoming), investment (Higgins 1998; Kopecky 2022), or in patterns of saving and consumption (Miles 1999; Borsch-Supan 2003).

[34] Note that this mean does *not* directly involve the demographics-constant rate. Rather, it involves the mean CPS unemployment rate constructed without any regression adjustment.

[35] A positive difference between the LAUS rate and the CPS mean rate in year t–1 suggests that a rate constructed with CPS microdata will be higher in year t if sampling error is responsible for the t–1 difference.

Third, the indirect effects could stem from the rate of labor force growth, which is correlated with the age distribution, rather than from the age distribution itself. In particular, a younger labor force is typically associated with a higher rate of labor force growth, and in the short run, labor demand may not fully keep up with this growth. If so, then higher unemployment would result. This notion is in the spirit of Cole and Rogerson (1999) and Elsby, Michaels, and Solon (2009), who show that contrary to the Mortensen-Pissarides model, job creation does not immediately increase in response to an increase in unemployment so as to quickly return unemployment to its new steady-state level.[36] Additionally, a sluggish response to labor force growth is consistent with the findings of Petrosky-Nadeau, Zhang, and Kuehn (2018), who argue that congestion externalities in the labor market amplify increases in unemployment. While these studies refer to cyclical frequencies, it is plausible that something similar is at work at lower frequencies. DellaVigna and Pollet (2007) find that investment in specific product lines is slow to respond to easily forecasted changes in demand that are driven by demographics, and Karahan, Pugsley, and Şahin (2021) relate the rate of business formation to the labor force growth caused by slowly moving demographic trends.

Evaluating the contributions of these various channels is beyond the scope of this paper, but it is a valuable area of future research.

8 Conclusion

In this paper, we revisit the influence of the age distribution of the population on the unemployment rate to go beyond the standard shift-share analysis. Expanding on previous work, we construct an index that captures the implications of the shift-share hypothesis. Regressing the unemployment rate on this index at the state level yields point estimates greater than one, implying that the effect of the age distribution on the unemployment rate is at least as strong as a shift-share method would predict, and indicating the presence of indirect effects that amplify the direct, mechanical influence of the age distribution on unemployment. IV estimation and spatial models reinforce this conclusion, as do regressions that relate the age distribution to unemployment rates within age groups.

This result should interest both policymakers and researchers who study the macroeconomics of the labor market. As the share of young persons in the US population has declined, estimates of the natural rate of unemployment have generally been revised downward over time, often based on shift-share methods. Our results suggest that the decline in the natural

[36]Similarly, Shimer (1998) argues that as the baby-boom generation entered the labor market, firms' demand for labor did not keep pace with the growing supply of labor. This discrepancy had the largest effect on young workers, who were new to the labor market, so the national youth unemployment rate rose.

rate may have been larger than these methods indicate. Additionally, estimates of the natural rate of unemployment were typically raised in the 1970s, when the demographic dynamic was working in the other direction. Several possibilities for the nature of indirect effects can be drawn from the literature. Evaluating the role of these mechanisms in generating our results may point toward a better understanding of the forces influencing unemployment.

References

Aaronson, Daniel, Luojia Hu, Arian Seifoddini, and Daniel G. Sullivan. 2015. "Changing Labor Force Composition and the Natural Rate of Unemployment." *Chicago Fed Letter* 2015(338). Available at https://www.chicagofed.org/publications/chicago-fed-letter/2015/338.

Acemoglu, Daron, and Pascual Restrepo. 2017. "Secular Stagnation? The Effect of Aging on Economic Growth in the Age of Automation." *American Economic Review: Papers & Proceedings* 107(5): 174–179. doi:10.1257/aer.p20171101.

Barnichon, Regis, and Geert Mesters. 2018. "On the Demographic Adjustment of Unemployment." *Review of Economics and Statistics* 100(2): 219–231. doi:10.1162/rest_a_00688.

Belotti, Federico, Gordon Hughes, and Andrea Piano Mortari. 2013. "XSMLE: Stata Module for Spatial Panel Data Models Estimation." Statistical Software Components S457610, Boston College Department of Economics, revised June 7, 2017. Available at http://fmwww.bc.edu/repec/bocode/x/xsmle.ado.

Belotti, Federico, Gordon Hughes, and Andrea Piano Mortari. 2017. "Spatial Panel-Data Models Using Stata." *Stata Journal* 17(1): 139–180. doi:10.1177/1536867X1701700109.

Borsch-Supan, Axel. 2003. "Labor Market Effects of Population Aging." *Labour* 17(Special Issue): 5–44. doi:10.1111/1467-9914.17.specialissue.2.

Bureau of Labor Statistics, US Department of Labor. 1974. "New Procedures For Estimating Unemployment In States and Local Areas." Report 432.

Cajner, Tomaz, and Isabel Cairo. 2013. "The Fading Dynamism of the US Labor Market: The Role of Demographics." 2013 Meeting Papers from Society for Economic Dynamics, No. 1208. Available at https://econpapers.repec.org/paper/redsed013/1208.htm.

Clark, Robert L., and Beth M. Ritter. 2020. "How Are Employers Responding to an Aging Workforce?" National Bureau of Economic Research Working Paper No. 26633.

Cole, Harold L., and Richard Rogerson. 1999. "Can the Mortensen-Pissarides Matching Model Match the Business-Cycle Facts?" *International Economic Review* 40(4): 933–959. doi:10.1111/1468-2354.00048.

Congressional Budget Office. 2019. "An Update to the Budget and Economic Outlook: 2019–2029." Available at: https://www.cbo.gov/system/files/2019-08/55551-CBO-outlook-update_0.pdf.

Cosic, Damir, and C. Eugene Steuerle. 2021. "Is Demand for Older Workers Adjusting to an Aging Labor Force?" doi:10.2139/ssrn.3961000. Center for Retirement Research at Boston College Working Paper 2021-18. Available at https://crr.bc.edu/working-papers/is-demand-for-older-workers-adjusting-to-an-aging-labor-force/.

Crump, Richard K., Marc Giannoni, Stefano Eusepi, and Ayşegül Şahin. 2019. "A Unified Approach to Measuring u*." *Brookings Papers on Economic Activity* 143–238. doi:10.1353/eca.2019.0002.

DellaVigna, Stefano, and Joshua M. Pollet. 2007. "Demographics and Industry Returns." *American Economic Review* 97(5): 1667–1702. doi:10.1257/aer.97.5.1667.

Driscoll, John C., and Aart C. Kraay. 1998. "Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data." *Review of Economics and Statistics* 80(4): 549–560. doi:10.1162/003465398557825.

Eggertsson, Gauti B., Manuel Lancastre, and Lawrence H. Summers. 2019. "Aging, Output Per Capita, and Secular Stagnation." *American Economic Review: Insights* 1(3): 325–342. doi:10.1257/aeri.20180383.

Elsby, Michael W. L., Ryan Michaels, and Gary Solon. 2009. "The Ins and Outs of Cyclical Unemployment." *American Economic Journal: Macroeconomics* 1(1): 84–110. doi:10.1257/mac.1.1.84.

Engbom, Niklas, et al. 2019. "Firm and Worker Dynamics in an Aging Labor Market." doi:10.21034/wp.756. Federal Reserve Bank of Minneapolis Working Paper 756. Available at https://www.minneapolisfed.org/research/working-papers/firm-and-worker-dynamics-in-an-aging-labor-market.

Fair, Ray C., and Kathryn M. Dominguez. 1991. "Effects of the Changing US Age Distribution on Macroeconomic Equations." *American Economic Review* 81(5): 1276–1294. Available at https://www.jstor.org/stable/2006917.

Foote, Christopher L. 2007. "Space and Time in Macroeconomic Panel Data: Young Workers and Unemployment Revisited." Federal Reserve Bank of Boston Working Paper No. 07-10. Available at https://www.bostonfed.org/-/media/Documents/Workingpapers/PDF/wp0710.pdf.

Glover, Andrew, and Jacob Short. 2020. "Demographic Origins of the Decline in Labor's Share." BIS Working Paper No. 874. Available at https://www.bis.org/publ/work874.pdf.

Hetze, Pascal, and Carsten Ochsen. 2006. "Age Effects on Equilibrium Unemployment." Manuscript, Rostock Center for the Study of Demographic Change, December. Available at https://www.rostockerzentrum.de/content/publikationen/rz_diskussionpapier_1.pdf.

Higgins, Matthew. 1998. "Demography, National Savings, and International Capital Flows." *International Economic Review* 39(2): 342–369. doi:10.2307/2527297.

Karahan, Fatih, Benjamin Pugsley, and Ayşegül Şahin. 2021. "Demographic Origins of the Startup Deficit." Federal Reserve Bank of New York Staff Report 888, available at https://www.newyorkfed.org/medialibrary/media/research/staff_reports/sr888.pdf.

Karahan, Fatih, and Serena Rhee. 2017. "Population Aging, Migration Spillovers, and the Decline in Interstate Migration." Federal Reserve Bank of New York Staff Report 699. Available at https://www.newyorkfed.org/medialibrary/media/research/staff_reports/sr699.pdf.

Kopecky, Joseph. 2019. "An Aging Dynamo: Demographic Change and the Decline of Entrepreneurial Activity in the United States." Manuscript. Available at https://www.josephkopecky.com/Papers/AgingEntrepreneurs.pdf.

Kopecky, Joseph. 2022. "Population Age Structure and Secular Stagnation: The Long Run Evidence." doi:10.2139/ssrn.4202217. Manuscript.

Korenman, Sanders, and David Neumark. 2000. "Cohort Crowding and Youth Labor Markets: A Cross-National Analysis." In *Youth Employment and Joblessness in Advanced Countries*, David Blanchflower and Richard Freeman, eds., 57–105. Chicago: University of Chicago Press.

Lee, Lung-fei, and Jihai Yu. 2010. "Estimation of Spatial Autoregressive Panel Data Models with Fixed Effects." *Journal of Econometrics* 154(2): 165–185. doi:10.1016/j.jeconom.2009.08.001.

Liang, James, Hui Wang, and Edward P. Lazear. 2018. "Demographics and Entrepreneurship." *Journal of Political Economy* 126(S1): S140–S196. doi:10.1086/698750.

Maestas, Nicole, Kathleen J. Mullen, and David Powell. forthcoming. "The Effect of Population Aging on Economic Growth, the Labor Force, and Productivity." *American Economic Journal: Macroeconomics*.

Mankiw, N. Gregory, and David N. Weil. 1989. "The Baby Boom, the Baby Bust, and the Housing Market." *Regional Science and Urban Economics* 19(2): 235–258. doi:10.1016/0166-0462(89)90005-7.

Miles, David. 1999. "Modelling the Impact of Demographic Change Upon the Economy." *Economic Journal* 109(452): 1–36. doi:10.1111/1468-0297.00389.

Neumark, David, and Maysen Yen. 2020. "Relative Sizes of Age Cohorts and Labor Force Participation of Older Workers." *Demography* 57(1): 1–31. doi:10.1007/s13524-019-00837-8.

Nickell, Stephen. 1981. "Biases in Dynamic Models with Fixed Effects." *Econometrica* 49(6): 1417–1426. doi:10.2307/1911408.

Norwood, Janet L. 1977. "Reshaping a Statistical Program To Meet Legislative Priorities." *Monthly Labor Review* 100(11): 6–11. Available at https://www.jstor.org/stable/41840357.

Ochsen, Carsten. 2021. "Age Cohort Effects on Unemployment in the USA: Evidence from the Regional Level." *Papers in Regional Science* 100(4): 1025–1053. doi:10.1111/pirs.12599.

Perry, George L. 1970. "Changing Labor Markets and Inflation." *Brookings Papers on Economic Activity* 1970(3): 411–448. doi:10.2307/2534139.

Petrosky-Nadeau, Nicolas, Lu Zhang, and Lars-Alexander Kuehn. 2018. "Endogenous Disasters." *American Economic Review* 108(8): 2212–45. doi:10.1257/aer.20130025.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Megan Schouweiler, and Matthew Sobek. 2022. "IPUMS USA: Version 12.0." doi:10.18128/D010.V12.0. Accessed July 27, 2022; retrieved from https://usa.ipums.org/usa/.

Shackleton, Robert. 2018. "Estimating and Projecting Potential Output Using CBO's Forecasting Growth Model." Congressional Budget Office Working Paper 2018-03. Available at https://www.cbo.gov/system/files/115th-congress-2017-2018/workingpaper/53558-cbosforecastinggrowthmodel-workingpaper.pdf.

Sheiner, Louise. 2014. "The Determinants of the Macroeconomic Implications of Aging." *American Economic Review* 104(5): 218–23. doi:10.1257/aer.104.5.218.

Shimer, Robert. 1998. "Why is the U.S. Unemployment Rate So Much Lower?" *NBER Macroeconomics Annual* 13: 11–74.

Shimer, Robert. 2001. "The Impact of Young Workers on the Aggregate Labor Market." *Quarterly Journal of Economics* 116(3): 969–1007. doi:10.1162/00335530152466287.

Skans, Oskar Nordstrom. 2005. "Age Effects in Swedish Local Labor Markets." *Economics Letters* 86(3): 419–426. doi:10.1016/j.econlet.2004.09.004.

Solon, Gary, Steven J. Haider, and Jeffrey M. Wooldridge. 2015. "What Are We Weighting For?" *Journal of Human Resources* 50(2): 301–316. doi:10.3368/jhr.50.2.301.

Surveillance Research Program of the National Cancer Institute. 2022. "Surveillance, Epidemiology, and End Results (SEER) Program Populations (1969-2019)." National Institutes of Health. Accessed July 12, 2022; retrieved from https://seer.cancer.gov/popdata/download.html.

Tuzemen, Didem. 2017. "A New Estimate of the Natural Rate of Unemployment." *Federal Reserve Bank of Kansas City Economic Bulletin* November 29. Available at https://www.kansascityfed.org/research/economic-bulletin/new-estimate-natural-rate-unemployment-2017/.

United States Mortality Database. 2021. University of California, Berkeley. Accessed July 25, 2022; retrieved from https://usa.mortality.org/.

Wachter, Michael L. 1977. "Intermediate Swings in Labor-Force Participation." *Brookings Papers on Economic Activity* 1977(2): 545–576.

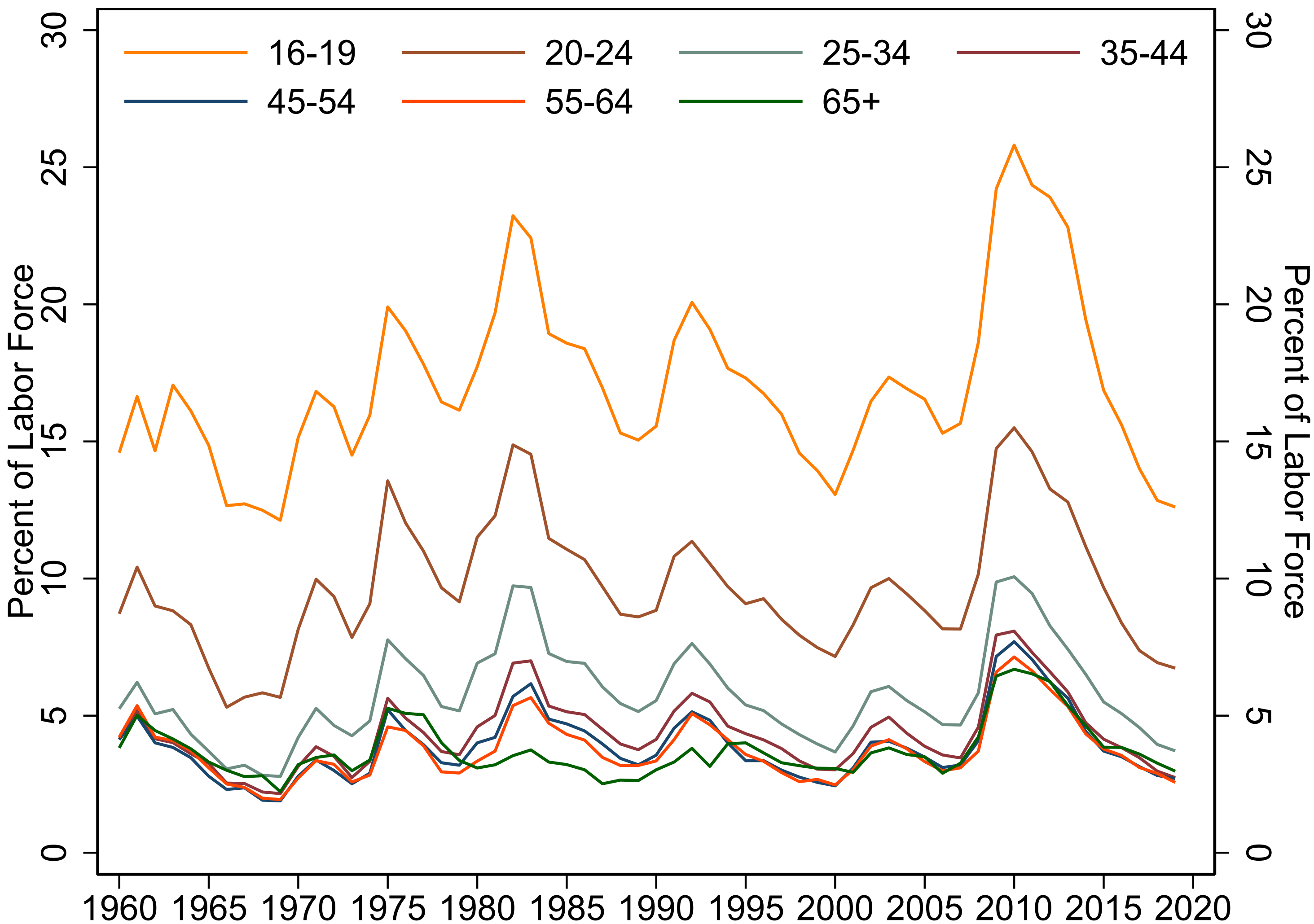

Figure 1. Unemployment Rates by Age Group: 1960–2019. *Source:* Bureau of Labor Statistics/Haver Analytics.

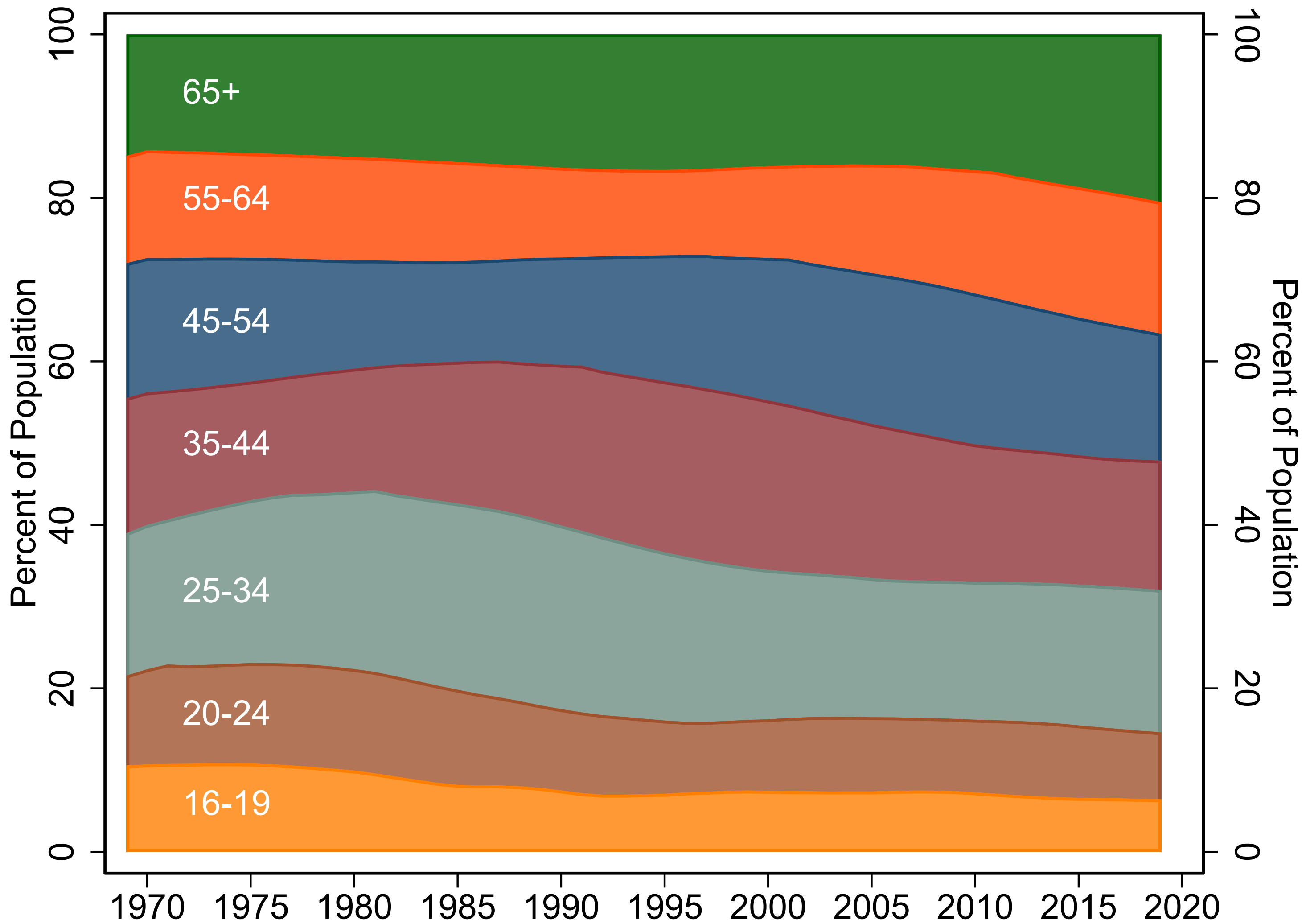

Figure 2. Population Shares of Age Groups: 1969–2019. *Source:* Surveillance, Epidemiology, and End Results Program (SEER, 2022).

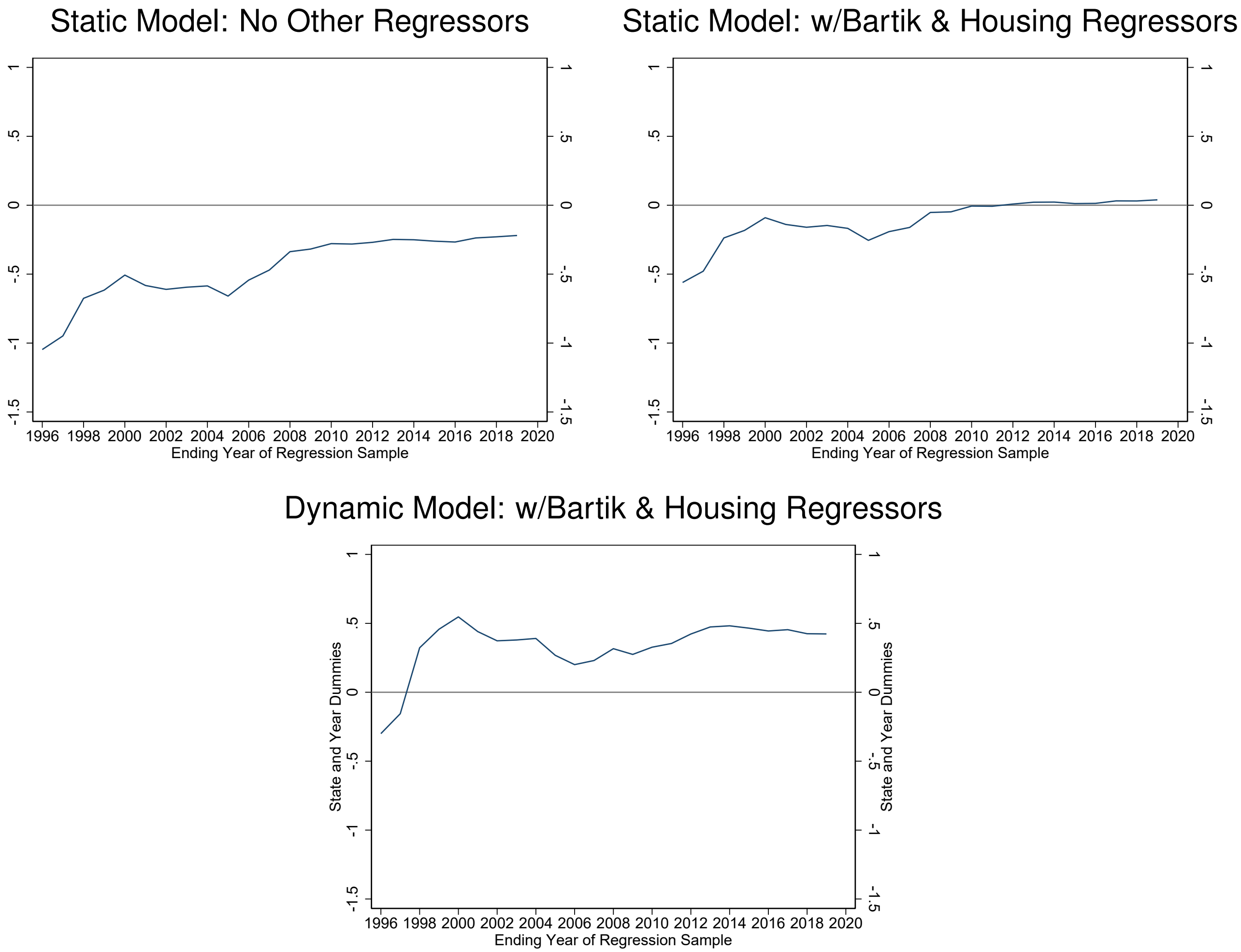

Figure 3. The Evolution of the Estimated Youth-Share Effect on State-Level Unemployment Rates. *Note:* Series correspond to point estimates of the youth-share coefficient (top panels) or long-run youth-share effect (bottom panel) from state-level regressions using samples of varying lengths. Ending dates of samples vary along the horizontal axes of the panels, but for each state, sample periods always begin in either 1970 or 1973, depending on the year that unofficial BLS unemployment rates become available for that state. Official BLS unemployment rates are used for all states from 1976 on. All regressions include state and year fixed effects, and both the unemployment rate and the youth share (defined as the share of persons aged 16 to 24 relative to the population aged 16 to 64) are in logs. The top two panels are estimated with Prais-Winsten AR(1) corrections, while the bottom panel uses a lagged dependent variable and OLS. The long-run youth-share effect in the bottom panel is calculated as the youth-share coefficient divided by one minus the estimated coefficient on the lagged unemployment rate. The top-right and bottom panels also include the Bartik and housing covariates (and two lags of each). The Bartik variables are weighted averages of national industry-specific employment-growth rates, with industry shares in state-level gross domestic product as the weights. Housing variables are log changes in total construction permits issued by the state, normalized by population. Samples exclude Alaska; Washington, DC; and Hawaii. *Source:* Authors' calculations.

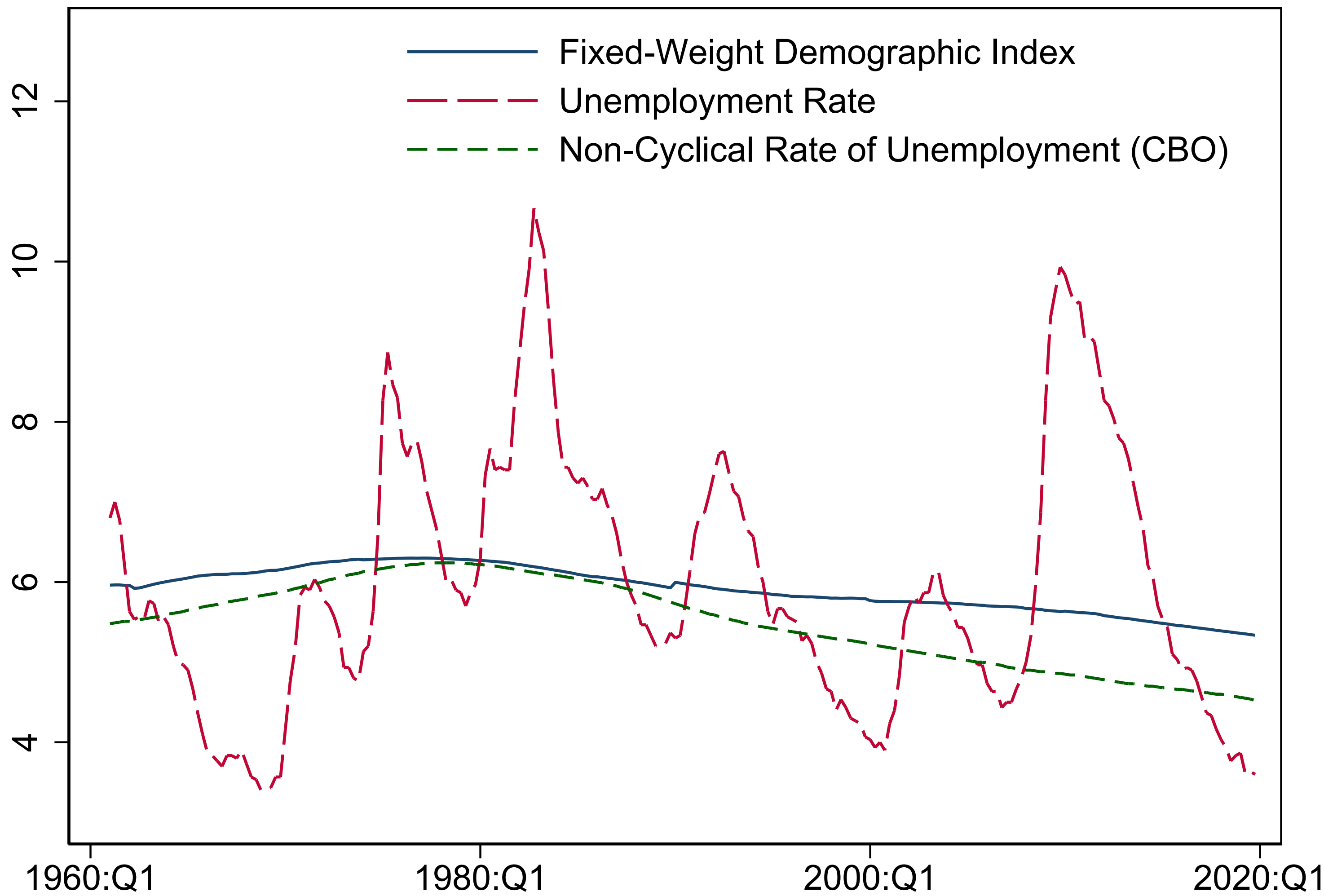

Figure 4. A Demographic Index for Unemployment, the Actual US Unemployment Rate, and the Non-Cyclical Rate of Unemployment: 1960:Q1 to 2019:Q4. *Note:* The demographic index is constructed by the authors and designed to reflect expected changes in unemployment if demographic shifts do not cause age-specific unemployment rates to change. It is a weighted average of time-varying population shares for comprehensive and mutually exclusive US age groups, with fixed weights that reflect average age-specific unemployment rates and labor force participation rates during the sample period. See the text and data appendix for details. The non-cyclical rate of unemployment is calculated by the Congressional Budget Office. It is often referred to as the CBO NAIRU. *Source:* Congressional Budget Office/Haver Analytics, Bureau of Labor Statistics/Haver Analytics, and authors' calculations.

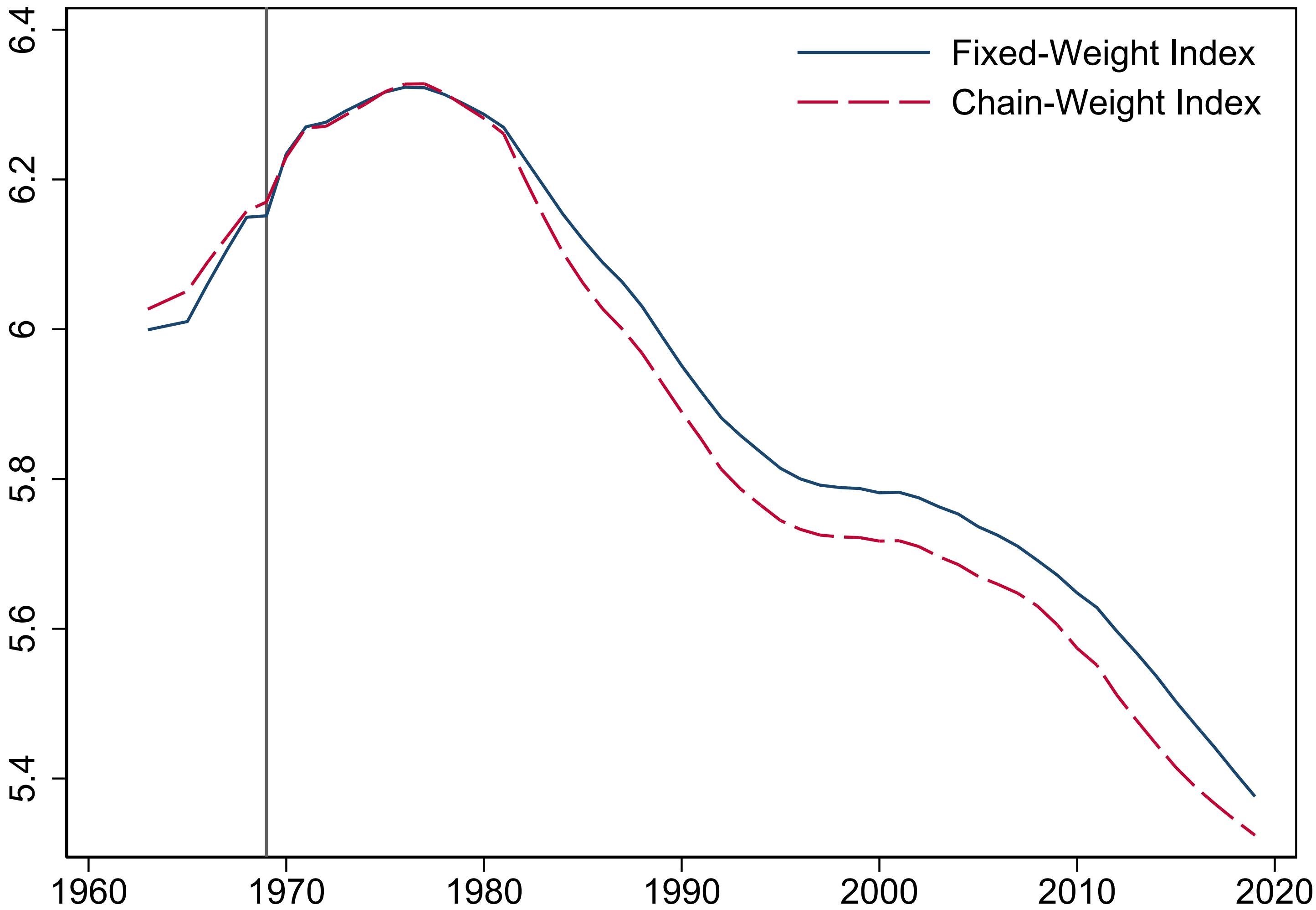

Figure 5. Fixed-Weight and Chain-Weight Demographic Indexes for the United States: 1963 through 2019. *Note:* These indexes reflect expected movements in national unemployment if the age distribution has no indirect effects on age-specific unemployment rates. The fixed-weight demographic index is a weighted average of state-level population shares for various age groups, with the (fixed) weights reflecting average age-specific unemployment rates and labor force participation rates. The chain-weight index allows these weights to change over time. The indexes omit data from Alaska; Washington, DC; and Hawaii. *Source:* Authors' calculations using data from the Bureau of Labor Statistics and the SEER population program.

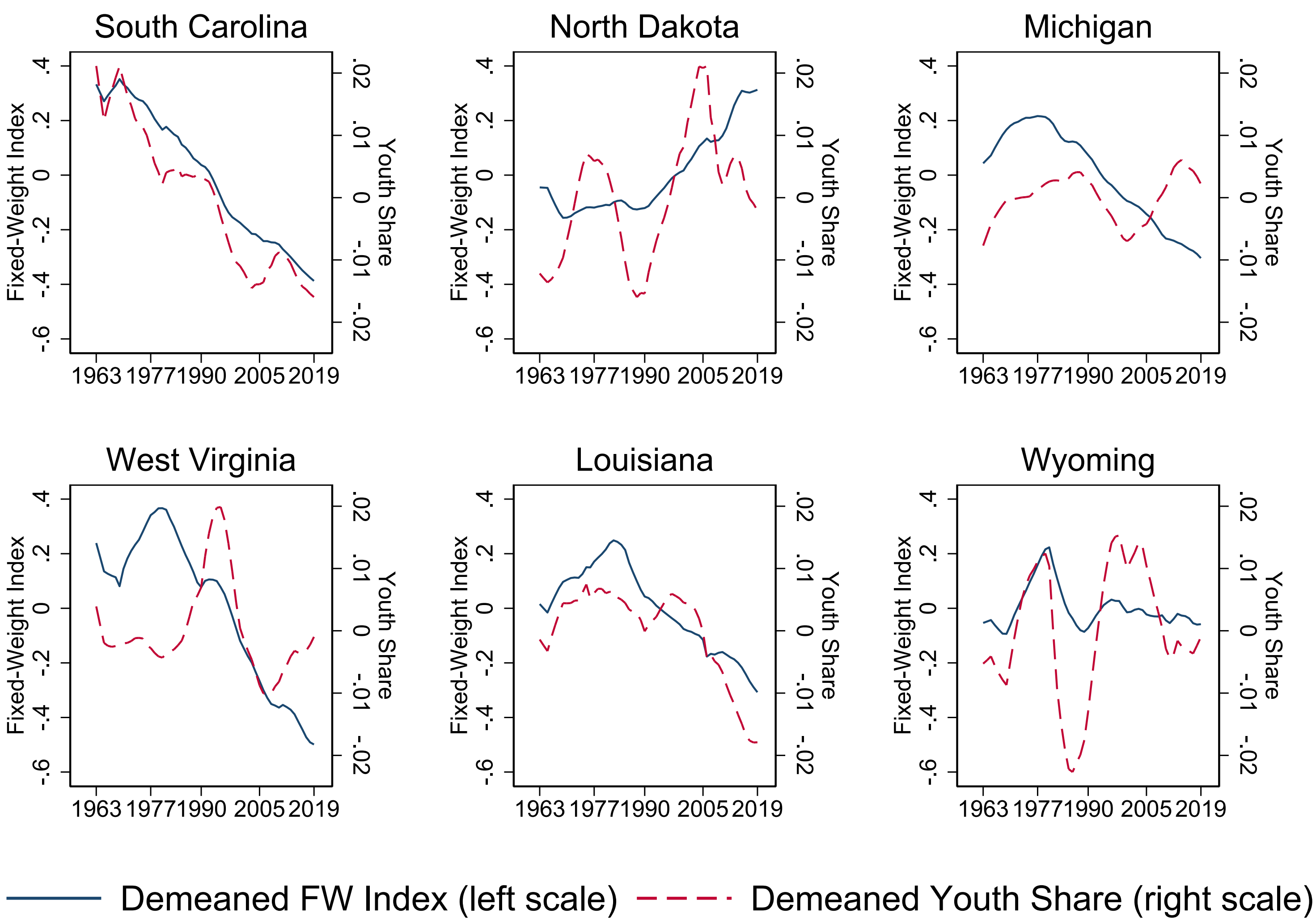

Figure 6. Fixed-Weight (FW) Demographic Indexes and Youth Shares of Population for Selected States: 1963 through 2019. *Note:* Blue lines correspond to fixed-weight demographic indexes that reflect the expected movements of the unemployment rate in a state if there are no indirect effects of demographic shifts on age-specific unemployment rates. Red dashed lines correspond to the share of 16- to 24-year-olds relative to the population of 16- to 64-year-olds. All series are demeaned from state and year averages. *Source:* Authors' calculations (demographic indexes) and SEER population program (youth shares).

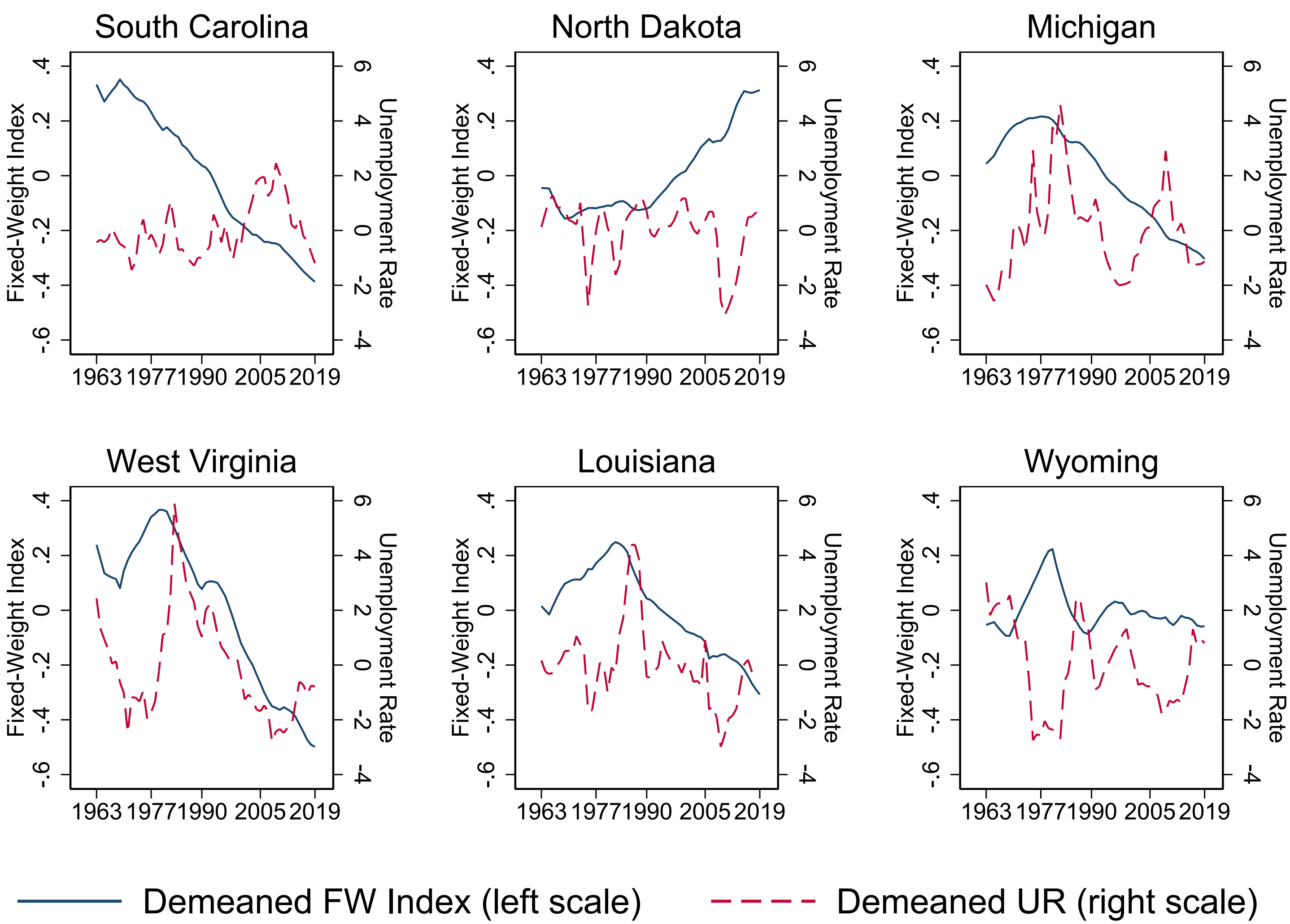

Figure 7. Fixed-Weight (FW) Demographic Indexes and Unemployment Rates for Selected States: 1963 through 2019. *Note:* Blue lines correspond to fixed-weight demographic indexes that reflect the expected movements of the unemployment rate in a state if there are no indirect effects of demographic shifts on age-specific unemployment rates. Red dashed lines are state-level unemployment rates. Unemployment rates prior to 1970 for all states and prior to 1973 for some states are inferred by the authors from unofficial state-level unemployment rates generated largely from state unemployment-insurance programs; see the data appendix for details. All series are demeaned from state and year averages. *Source:* Authors' calculations (demographic indexes) and Bureau of Labor Statistics and authors' calculations (unemployment rates).

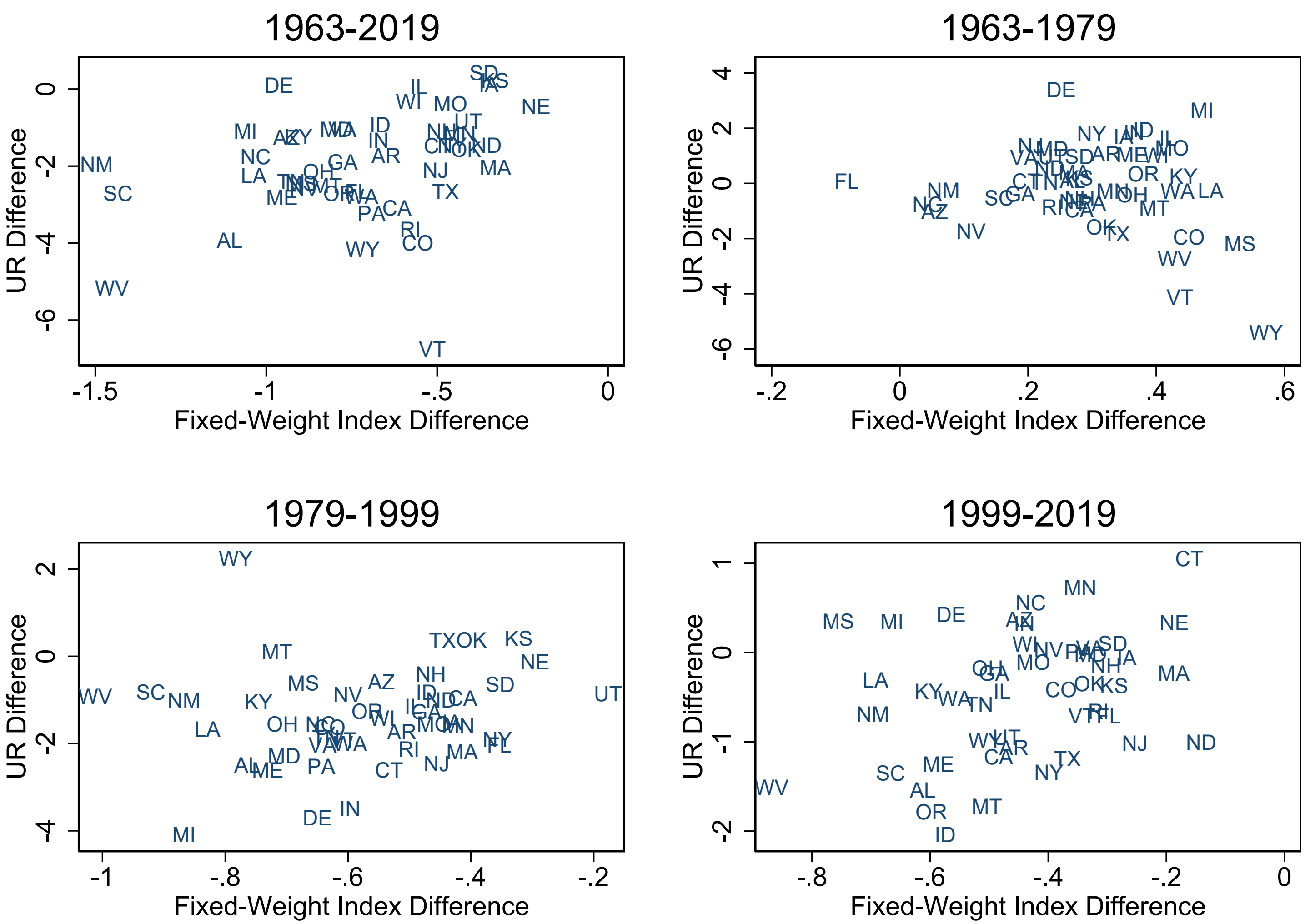

Figure 8. Long Differences of State-Level Fixed-Weight Demographic Indexes and Unemployment Rates: 1963 through 2019. *Note:* Unemployment rates prior to 1970 for all states and prior to 1973 for some states are inferred by the authors from unofficial state-level unemployment rates generated largely from state unemployment-insurance programs; see the data appendix for details. Samples exclude Alaska; Washington, DC; and Hawaii. *Source:* Authors' calculations (demographic indexes) and Bureau of Labor Statistics and authors' calculations (unemployment rates).

	1963-2019		1963-1979		1979-1999		1999-2019	
Ln Pop-Weighted?	No	Yes	No	Yes	No	Yes	No	Yes
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Change in Fixed-Weight Index	1.613*	1.610*	-2.014	-1.258	1.066	1.238	1.430*	1.390*
	(0.654)	(0.641)	(2.232)	(1.699)	(1.030)	(0.904)	(0.674)	(0.624)
Constant	-0.798	-0.784	0.532	0.348	-0.798	-0.744	0.169	0.161
	(0.531)	(0.494)	(0.585)	(0.556)	(0.541)	(0.537)	(0.297)	(0.294)
N	48	48	48	48	48	48	48	48

Table 1. Long-Difference Regressions. *Note:* $^{+}\ p < 0.10$, $^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$. The dependent variable in each regression is the change in the state's unemployment rate during the time period specified. Regressions in even-numbered columns are weighted by the natural log of the state's mean population from 1963 through 2019. Robust standard errors in parentheses. Samples exclude Alaska; Washington, DC; and Hawaii. *Source:* Authors' calculations.

Ln Pop-Weighted?	No (1)	No (2)	No (3)	Yes (4)	No (5)	No (6)	No (7)	Yes (8)
Lagged Dep. Var.	0.845***	0.823***	0.823***	0.822***	0.847***	0.824***	0.825***	0.823***
	(0.0234)	(0.0225)	(0.0220)	(0.0209)	(0.0231)	(0.0220)	(0.0215)	(0.0204)
Fixed-Weight Index	0.427*	0.422+	0.399+	0.395+				
	(0.199)	(0.226)	(0.209)	(0.217)				
Chain-Weight Index					0.446*	0.432+	0.385+	0.381+
					(0.209)	(0.222)	(0.198)	(0.204)
Bartik		-33.76***	-28.55***	-29.35***		-33.71***	-28.55***	-29.34***
		(3.838)	(3.215)	(3.273)		(3.862)	(3.248)	(3.310)
Bartik $t-1$		-0.762	2.505	2.327		-0.746	2.509	2.337
		(2.853)	(2.682)	(2.824)		(2.855)	(2.689)	(2.832)
Bartik $t-2$		3.979	5.971+	6.175+		3.922	5.924	6.132
		(4.044)	(3.560)	(3.690)		(4.085)	(3.605)	(3.733)
Housing			-0.789***	-0.793***			-0.784***	-0.788***
			(0.104)	(0.102)			(0.103)	(0.100)
Housing $t-1$			-0.532***	-0.548***			-0.528***	-0.545***
			(0.0832)	(0.0816)			(0.0818)	(0.0800)
Housing $t-2$			-0.190*	-0.193*			-0.187*	-0.190*
			(0.0880)	(0.0914)			(0.0862)	(0.0897)
Long-run Effect	**2.76**	**2.38**	**2.25**	**2.22**	**2.92**	**2.46**	**2.20**	**2.16**
LR Effect Std. Err.	**1.37**	**1.26**	**1.13**	**1.17**	**1.48**	**1.24**	**1.10**	**1.12**
No. of States	48	48	48	48	48	48	48	48
N	2736	2736	2736	2736	2736	2736	2736	2736

Table 2. OLS Unemployment Regressions: 1963 through 2019. *Note:* $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The dependent variable in each regression is the state's unemployment rate. The fixed-weight demographic index is a weighted average of state-level population shares for various age groups, with the (fixed) weights reflecting average age-specific unemployment rates and labor force participation rates. The chain-weight index allows these weights to change over time. The long-run effect in bold is calculated as the demographic-index coefficient estimate divided by one minus the estimated coefficient for the lagged unemployment rate. The Bartik variables are weighted averages of national industry-specific employment-growth rates, with industry shares in state-level gross domestic product as the weights. Housing variables are log changes in total construction permits issued by the state, normalized by population. Standard errors are calculated using the Driscoll-Kraay method with a lag length of five years. Samples exclude Alaska; Washington, DC; and Hawaii. *Source:* Authors' calculations.

Method	OLS	IV	IV	OLS	IV	IV
Ln Pop-Weighted?	No	No	Yes	No	No	Yes
	(1)	(2)	(3)	(4)	(5)	(6)
Lagged Dep. Var.	0.795^{***}	0.796^{***}	0.796^{***}	0.795^{***}	0.796^{***}	0.796^{***}
	(0.0153)	(0.0152)	(0.0145)	(0.0145)	(0.0144)	(0.0138)
Fixed-Weight Index	0.704^{*}	0.643^{*}	0.633^{*}			
	(0.285)	(0.288)	(0.298)			
Chain-Weight Index				0.855^{**}	0.823^{**}	0.813^{**}
				(0.265)	(0.263)	(0.271)
First-stage F-stat		974.6	888.1		1188.2	996.8
Long-Run Effect	**3.43**	**3.16**	**3.11**	**4.18**	**4.03**	**3.98**
LR Effect Std. Err.	**1.36**	**1.39**	**1.45**	**1.25**	**1.26**	**1.30**
No. of States	48	48	48	48	48	48
N	1968	1968	1968	1968	1968	1968

Table 3. OLS and IV Unemployment Regressions: 1979 through 2019. *Note:* $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The dependent variable in each regression is the state's unemployment rate. Regressions also include the six Bartik and housing variables reported in the OLS regressions of Table 2. The instrumental variable used in columns 2, 3, 5, and 6 is a predicted demographic index based on the state's demographic makeup 10 years earlier. Calculation of this instrument uses either a fixed-weight formula or a chain-weight formula as appropriate. The long-run effect of the demographic index in bold is calculated as the demographic-index coefficient estimate divided by one minus the estimated coefficient for the lagged unemployment rate. Standard errors are calculated using the Driscoll-Kraay method with a lag length of five years. Samples exclude Alaska; Washington, DC; and Hawaii. *Source:* Authors' calculations.

Method	OLS (1)	1963–2019 Spatial (2)	Spatial (3)	OLS (4)	1979–2019 Spatial (5)	Spatial (6)
Lagged Dep. Var.	0.823^{***} (0.0220)	0.791^{***} (0.0318)	0.788^{***} (0.0314)	0.795^{***} (0.0153)	0.759^{***} (0.0186)	0.758^{***} (0.0190)
Fixed-Weight Index	0.399^{+} (0.209)	0.504^{**} (0.181)	0.544^{**} (0.166)	0.704^{*} (0.285)	0.758^{**} (0.246)	0.765^{**} (0.242)
FW Index: Spatial Lag			-0.316 (0.216)			-0.0597 (0.149)
Spatial Autoregressive Coeff.		0.114^{*} (0.0453)	0.120^{**} (0.0450)		0.139^{***} (0.0272)	0.140^{***} (0.0310)
Spatial Error Coeff.		0.218^{**} (0.0821)	0.212^{**} (0.0811)		0.196^{***} (0.0460)	0.195^{***} (0.0476)
SSE		0.212^{***} (0.0395)	0.212^{***} (0.0394)		0.172^{***} (0.0251)	0.172^{***} (0.0251)
Long-Run Effect	**2.25**	**2.42**	**2.57**	**3.43**	**3.14**	**3.16**
LR Effect Std. Err.	**1.13**	**0.85**	**0.78**	**1.36**	**1.01**	**0.99**
P-val: LR Eff = 0	0.047	0.005	0.001	0.011	0.002	0.001
P-val: LR Eff = 1	0.269	0.098	0.045	0.073	0.034	0.029
N	2736	2736	2736	1968	1968	1968

Table 4. OLS and Spatial-Panel Unemployment Regressions: 1963 through 2019 and 1979 through 2019. *Note:* $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The dependent variable in each regression is the state's unemployment rate. Regressions also include the six Bartik and housing variables reported in the OLS regressions of Table 2. The spatial-panel regressions in columns 2, 3, 5, and 6 use a row-normalized first-order contiguity matrix for both the spatial autoregressive component and the spatial error component. The long-run effect of the demographic index in bold is calculated as the demographic-index coefficient estimate divided by one minus the estimated coefficient for the lagged unemployment rate. The spatial lag of the fixed-weight index in the third row is a weighted average of year-t fixed-weight indexes in bordering states, using the weights in the normalized first-order contiguity matrix. Standard errors are calculated using the Driscoll-Kraay method with a lag length of five years. Samples exclude Alaska; Washington, DC; and Hawaii. *Source:* Authors' calculations.

Panel A: OLS and Spatial-Panel Regressions

	Young (16-24)		Prime (25-54)		Older (55+)	
	OLS	Spatial	OLS	Spatial	OLS	Spatial
	(1)	(2)	(3)	(4)	(5)	(6)
Lagged Dep. Var.	0.558***	0.501***	0.641***	0.546***	0.364***	0.329***
	(0.0395)	(0.0271)	(0.0327)	(0.0212)	(0.0192)	(0.0166)
Ln FW Index	0.181	0.265	0.414	0.521**	0.353	0.398
	(0.260)	(0.177)	(0.257)	(0.181)	(0.576)	(0.359)
Spatial Autoregressive Coeff.		0.297***		0.310***		0.366***
		(0.0461)		(0.0236)		(0.0549)
Spatial Error Coeff.		-0.265***		-0.151**		-0.298***
		(0.0383)		(0.0460)		(0.0844)
SEE		0.0138***		0.0116***		0.0480***
		(0.000960)		(0.000513)		(0.00643)
N	1968	1968	1968	1968	1968	1968

Panel B: IV Regressions

	Young (16-24)	Prime (25-54)	Older (55+)
	IV	IV	IV
	(1)	(2)	(3)
Lagged Dep. Var.	0.561***	0.641***	0.364***
	(0.0400)	(0.0328)	(0.0196)
Ln FW Index	0.809***	0.484$^+$	0.749
	(0.232)	(0.268)	(0.559)
N	1968	1968	1968

Table 5. OLS, Spatial-Panel, and IV Regressions for Age-Specific Unemployment Rates: 1979 through 2019. *Note:* $^+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The dependent variable in each regression is the natural log of the unemployment rate for the given age group in a state. Regressions also include the six Bartik and housing variables reported in the OLS regressions of Table 2. Regressions in even-numbered columns in the upper panel use a spatial-panel method, with a row-normalized first-order contiguity matrix used for both the spatial autoregressive component and the spatial error component. The instrumental variable used in the lower panel is the predicted demographic index for the state based on its demographic makeup 10 years earlier. Standard errors are calculated using the Driscoll-Kraay method with a lag length of five years. Samples exclude Alaska; Washington, DC; and Hawaii. *Source:* Authors' calculations.

	OLS (1)	Spatial (2)	OLS (3)	Spatial (4)	OLS (5)	Spatial (6)	OLS (7)	Spatial (8)
Lagged Dep. Var.	0.723***	0.640***	0.717***	0.633***	0.749***	0.681***	0.741***	0.674***
	(0.0310)	(0.0236)	(0.0327)	(0.0231)	(0.0300)	(0.0226)	(0.0320)	(0.0237)
Ln FW Index	0.279	0.397*			0.433+	0.526**		
	(0.238)	(0.175)			(0.226)	(0.191)		
Ln Predicted FW Index			0.351+	0.440**			0.398*	0.471**
			(0.211)	(0.162)			(0.195)	(0.174)
Lagged LAUS–CPS Difference					0.768***	0.736***	0.760***	0.728***
					(0.0719)	(0.0779)	(0.0727)	(0.0787)
Spatial Autoregressive Coeff.		0.241***		0.240***		0.194***		0.191***
		(0.0267)		(0.0260)		(0.0294)		(0.0287)
Spatial Error Coeff.		-0.0774		-0.0759		0.00181		0.00532
		(0.0708)		(0.0704)		(0.0671)		(0.0672)
SEE		0.00696***		0.00695***		0.00614***		0.00615***
		(0.000397)		(0.000401)		(0.000402)		(0.000409)
N	1968	1968	1968	1968	1968	1968	1968	1968

Table 6. OLS and Spatial-Panel Regressions for Demographics-Constant Unemployment Rates: 1979 through 2019. *Note:* + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The dependent variable in each regression is the natural log of a demographics-constant unemployment rate constructed at the state-year level using CPS microdata. These data are used to estimate yearly national unemployment logits, with regressors reflecting each individual's age group, gender, education, and race, along with state fixed effects. Coefficient estimates from these regressions are then multiplied times constant demographic weights to generate a demographics-constant unemployment rate for each state and year. The demographic weights do not vary across states or years, so variation in the rates across states within a year is generated solely by differing estimates of the state fixed effects across states. Variation in the demographics-constant unemployment rates across years is generated by changing state fixed effects and by changes in the estimated national demographic coefficients. The predicted fixed-weight index in the third row is the predicted demographic index for the state based on its demographic makeup 10 years earlier. The LAUS-CPS difference in the fourth row is the difference between the state's unemployment rate as published by the Local Area Unemployment Statistics (LAUS) program of the Bureau of Labor Statistics and the state unemployment rate generated from CPS microdata. Regressions also include the six Bartik and housing variables reported in the OLS regressions of Table 2. Regressions in even-numbered columns use a spatial-panel method, with a row-normalized first-order contiguity matrix used for both the spatial autoregressive component and the spatial error component. Standard errors are calculated using the Driscoll-Kraay method with a lag length of five years. Samples exclude Alaska; Washington, DC; and Hawaii. *Source:* Authors' calculations.

A Data Appendix

We compiled information from a variety of sources to construct the data needed for this paper.[1] This appendix describes those sources as well as our processing methods. Often, we needed to impute or interpolate data at low levels of aggregation (such as state or state-age groups) because official data were available only at higher levels of aggregation (such as nation). Where appropriate, we show comparisons between official data sources and our imputed estimates after aggregating them up to a higher level.

A.1 Population

A.1.1 1969–2019 data

Population data dating back to 1969 are available by county, age, race, and year from the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) program (Surveillance Research Program of the National Cancer Institute 2022). To construct data by state, age group, and year we simply aggregated population counts up from the county level. The data from SEER are top-coded at age 85 (that is, everyone aged 85 and older is collected into an 85-plus group). This has no effect on our final data, as the oldest age group with which we work is a 65-plus group.

A.1.2 Pre-1969 Population

Because we construct demographic indexes back to 1963, we also need population data for years that SEER does not cover. To estimate population for the 1960s, we interpolate state-level population by age and year between three years: 1960, 1965, and 1969. We obtain population counts for 1960 from the 5 percent sample of the 1960 US Census available from IPUMS (Ruggles et al. 2022). To estimate the population by state and age group for 1965, we make use of a question asked on the 1970 US Census regarding migration.[2] Specifically, this question asked where individuals lived five years earlier (that is, in 1965). However, persons answering this question in 1970 do not account for all people who lived in a particular state in 1965, because some of the 1965 residents moved out of the United States or died.

We assume that international out-migration accounts for an insignificant fraction of people who were not sampled in 1970, and that accounting for mortality alone is sufficient to represent population shares of certain age groups by state and year in 1965. To account

[1] Special thanks are due to Luke Stewart for his help in preparing this appendix.

[2] IPUMS features two different 1 percent samples from the 1970 US Census, one drawn from the original 5 percent sample and one from the original 15 percent sample. These samples can be combined with minimal complication, effectively yielding a 2 percent sample. See https://forum.ipums.org/t/questions-about-1970-census-data/3133 for details.

for mortality, we use the death rates compiled by age, state, gender, and year by the US Mortality Database at the University of California, Berkeley.[3] We use the death rates to back out an estimate of population in 1965 by age, state, and gender as follows. Let l_{ast} be the age-specific survival probability for residents of state s in year t, which is the probability of living from age a to $a+1$, starting in year t. Note that a refers to a single year of age rather than an age group, and that l_{ast} is equal to one minus the appropriate death rate. If $P^{1965}_{as,1965}$ is the actual population for a state and age group in 1965, then the population as measured by the 1970 US Census, $P^{1970}_{as,1965}$ will be

$$\begin{aligned} P^{1970}_{as,1965} &= P^{1965}_{as,1965} \cdot l_{as,1965} \cdot l_{a+1,s,1966} \cdot l_{a+2,s,1967} \cdot l_{a+3,s,1968} \cdot l_{a+4,s,1969} \\ &= P^{1965}_{as,1965} \prod_{\tau=0}^{4} l_{a+\tau,s,1965+\tau}. \end{aligned}$$

The actual population in 1965 can then be calculated as

$$P^{1965}_{as,1965} = \frac{P^{1970}_{as,1965}}{\prod_{\tau=0}^{4} l_{a+\tau,s,1965+\tau}}.$$

In short, by chaining survival rates together, we obtain a "population inflator" that improves the population counts by age and state in 1965, as opposed to simply using the information from the migration question in the 1970 US Census with no adjustment. For simplicity, the above expressions do not account for gender, but we do have both US Census data and mortality rates by state, age, year, and gender. We therefore perform the analysis separately for men and women because mortality rates differ by gender. An additional wrinkle is that to properly account for mortality among older individuals, we must estimate populations of individuals aged 85 and older due to the 85-plus top code in the SEER data. To do this, we chain together a product of probabilities of living from one age to the next in a manner similar to the method described above.

With population data from 1960 and 1965 in hand, we linearly interpolate population by age, state, and gender between the years 1960, 1965, and 1969 (the 1969 data come from SEER).[4] Figure A.1 shows a comparison between official US Census estimates of total state-level populations in 1965 to state-level aggregations of our inferred population counts for that year, adjusting and not adjusting for mortality. Comparing the bottom row of panels with the top row indicates that accounting for mortality notably improves our estimates. Figure A.2 compares official state-level population estimates in six selected states from 1960 through 1969 with our interpolated estimates aggregated to the state level. Each measure is

[3] See https://usa.mortality.org.

[4] Because there are three points through which to interpolate, a quadratic interpolation is possible, but we found that a quadratic did not considerably improve upon a linear model.

normalized to the official 1960 estimate. Although the interpolation works better for some states than others, we would caution that the official yearly US Census estimates of state-level population are also subject to error and were not constructed with the benefit of the retrospective information in the 1970 US Census.

Combining the pre-1969 estimates with the 1969–2019 SEER data, we have a series of populations by state, age, and year, which we aggregate up to the age-group level.

A.2 Unemployment Rates and Labor Force Participation Rates

We need unemployment rates and labor force participation rates by state, year, and age group for two reasons. First, both labor market variables are used to construct our fixed-weight and chain-weight demographic indexes. Second, we use age-group-specific unemployment rates on the left-hand side of some regressions.

A.2.1 1978–2019 data

The BLS publishes estimates of the population, labor force, employment, unemployment, unemployment rate, and labor force participation rate starting in 1978 by year, state, and age groups that match our age-group definitions. Therefore, we use the BLS estimates for the years 1978 through 2019.

A.2.2 Early State-Level Unemployment Rates

The BLS publishes official state-level unemployment rates from 1976 onward. State-level unemployment rates for earlier years can be difficult to locate.[5] In the 1960s, the Department of Labor published estimates of state-level unemployment rates in the *Manpower Report of the President*. These rates were constructed using a 70-step method intended to make unemployment rates based on data from unemployment-insurance systems comparable to unemployment rates constructed with CPS data. This method was described in the Bureau of Employment Security's *Handbook on Estimating Unemployment* and was known informally as the "handbook method."

Beginning in 1970 for large states and metropolitan statistical areas (MSAs), the BLS began to compare the handbook-method rates to mean state-level unemployment rates calculated with CPS data. By 1974, the BLS was benchmarking the handbook rates data to the CPS data for all states on a yearly basis. Essentially, the handbook rates would be adjusted arithmetically to the CPS rates once a year, and the monthly data for the rest of the year would be based on month-to-month differences between handbook estimates. Additionally, at the end of the year, handbook estimates would be re-estimated so that

[5] This discussion draws on Norwood (1977) and Bureau of Labor Statistics (1974).

their averages matched the CPS yearly average. This adjustment method suggests a way to adjust early-year handbook rates published in the *Manpower Reports* to later CPS-based estimates. Specifically, we adjust the *level* of the handbook-based estimates while preserving the year-to-year changes.

Based on the early CPS benchmarking efforts for large states, as well as the 1974 change in methodology for all states to CPS benchmarking, we obtained from the BLS and the *Manpower Reports* some unemployment rates calculated on both the handbook and directly from annual averages of CPS data. This overlap allowed us to calculate a level adjustment for the early handbook-based rates. Pre-1976 BLS unemployment rates that used the benchmarking method were supplied by Robert Shimer, who originally obtained them from Susan Gorel at the BLS. As a result, we sometimes refer to these data as the Shimer-Gorel data.

Let τ_s be the first year of overlap between the two series for state s (so that τ_s is the first year that CPS-based unemployment rates are recorded for state s). Let UR_{hst} be the handbook-based rate, UR_{cst} be the CPS-based rate, and UR_{st} be our standardized unemployment rate for state s in year t. Then

$$UR_{st} = UR_{hst} + (UR_{cs\tau} - UR_{hs\tau}).$$

That is, we use an adjustment factor based on the difference between the CPS-based rates and the handbook-based rates in year τ. Figure A.3 demonstrates the procedure for the state of Alabama, for which the CPS-benchmarked data began in 1970. The early BLS data consist of unpublished estimates, whereas official BLS estimates begin in 1976. Figure A.4 compares the aggregated unemployment rate calculated using this method with the official national unemployment rate from 1960 through 1975.

Critically, the soundness of this adjustment is largely determined by its application. Because we include the lag of the unemployment rate on the right-hand side of our models, and because these lagged dependent variables typically enter with coefficients near one, we are mostly concerned with the quality and consistency of year-over-year differences in state-level unemployment rates. Therefore, to the extent that the differences in the handbook-based rates align with the differences in the CPS-based rates, we have more confidence in using this method of adjustment. Figure A.5 plots differences between handbook-based rates and CPS-based rates when there is overlap between the series. This plot shows that the differences are positively correlated, lending credence to our methodology.

A.2.3 Pre-1978 Year-State-Age Group URs and LFPRs

As noted above, we need unemployment rates (URs) by year, state, and age group to construct our demographic indexes for years before 1978. However, these rates are not published by the BLS and cannot be constructed from CPS microdata, which begin in 1978. To estimate these rates for years before 1978, we therefore use post-1978 data and a regression model. The BLS publishes national data on unemployment rates by year, age group, and gender dating back to 1948. We fit the following model:

$$\ln UR_{ast} = \alpha + \beta^1_{as} \ln UR^{US}_{at} + \beta^2_{as} \ln UR^{US}_{t} + \varepsilon_{ast},$$

where UR_{ast} is the unemployment rate of age-gender group a in state s at time t. Note that we allow the coefficients to vary by age a and state s. Using OLS, we estimate this regression on the post-1978 data, using estimates from the monthly CPS (aggregated to the yearly level) to construct UR_{ast}.[6] By using nation-level data for demographic groups defined by age and gender to identify $\hat{\beta}_{as1}$, we capture particular factors that influence unemployment for the given demographic group within each state. By using the national aggregate unemployment rate to identify $\hat{\beta}_{as2}$, we also condition for the effect of the aggregate business cycle on the unemployment rate of the state-age-gender group.

We then use our estimates $\hat{\beta}_{as1}$ and $\hat{\beta}_{as2}$ to construct $\widehat{\ln UR_{ast}}$ for the pre-1978 sample. Figure A.6 compares those estimated unemployment rates, aggregated to the state level, with state-level unemployment rates described in the previous subsection. The vertical red line indicates the year 1978, before which the prediction is out of sample.

Our demographic indexes also require labor force participation rates (LFPRs) by year, state, age group, and gender. To construct these rates for the pre-1978 period, we use a simpler differences-in-means approach. For each age-gender group and state, we use post-1978 data to estimate:

$$\bar{\delta}_{as} = \frac{1}{T} \sum_{t=1}^{T} \left(\ln LFPR^{US}_{at} - \ln LFPR_{ast} \right),$$

which is the average difference between the LFPR of age-gender group a in state s and the LFPR of age-gender group a for the entire country. We then apply the estimate of $\bar{\delta}_{as}$ to calculate

$$\ln \widehat{LFPR}_{ast} = \ln LFPR^{US}_{at} + \bar{\delta}_{as}$$

for years before 1978. We found that this simpler approach led to better results in the case of LFPRs than a regression-based approach.

[6]CPS samples are available from IPUMS starting with 1976; samples for all states are available starting with 1978.

A.3 Bartik Variables

Our Bartik variables weight national industry-specific employment growth rates with industry shares of state-level gross state product (GSP). Specifically, we have

$$B_{st} = \sum_{j=1}^{J} w_{jst}(GR_{jt} - GR_t),$$

where j denotes industry, J is the total number of industries, w_{jst} is a GSP-based weight for industry j in state s in year t, GR_{jt} is the growth rate of national employment in industry j in year t, and GR_t is the growth rate of total national employment in year t. For the years prior to 1998, Standard Industry Classification (SIC) codes are used to differentiate industries, generally at the two-digit level. For the years since 1998, North American Industry Classification System (NAICS) codes are used, generally at the three-digit level.

As noted above, we use state-level industry output data to calculate the weights w_{jst}. Specifically, w_{jst} is given by

$$w_{jst} = \frac{Y_{js,t-1}}{Y_{s,t-1}},$$

where Y_{jst} is output in industry j in state s in year t, and Y_{st} is gross state product of state s in year t. That is, the weight of industry j in state s in year t is given by the share of total state output that was accounted for by j in year $t-1$. For both SIC-based and NAICS-based years, output data come from the Bureau of Economic Analysis (BEA) and include all private-industry classifications as well as government output (federal, state, and local).

Output data from the BEA are available only from 1963 onward, which allows us to construct weights only for 1964 onward. To extend our Bartiks back to 1961 to cover our full sample, we assume that industry shares do not change significantly at the state level from year to year. Therefore, we extend the 1963 industry shares back to 1960, and by doing so, we extend the 1964 weights back to 1961.

The formula for constructing the nation-level growth rates by industry is

$$GR_{jt} - GR_t = \frac{N_{jt} - N_{j,t-1}}{N_{j,t-1}} - \frac{N_t - N_{t-1}}{N_{t-1}},$$

where N_{jt} is now employment in industry j in year t. Employment data are available back to 1960. Therefore, using the expanded output weights, we are able to construct employment-based Bartiks for our full sample. Employment data come from the Bureau of Economic Analysis and are given in full-time equivalent (FTE) employment.

A.4 Housing Permit Index

We construct a state-year series reflecting the intensity of growth of housing to control for housing demand in our models. Data on new housing building permits come from the US Census Building Permits Survey. We construct

$$\Delta \ln C_{st} = \ln\left(\frac{c_{st}}{P_{st}}\right) - \ln\left(\frac{c_{s,t-1}}{P_{s,t-1}}\right),$$

where c_{st} is the number of new building permits issued in state s at time t, and P_{st} gives the population in state s at time t. Using the Taylor expansion of the natural log, we are using an approximation of the growth rate of new housing units per capita.

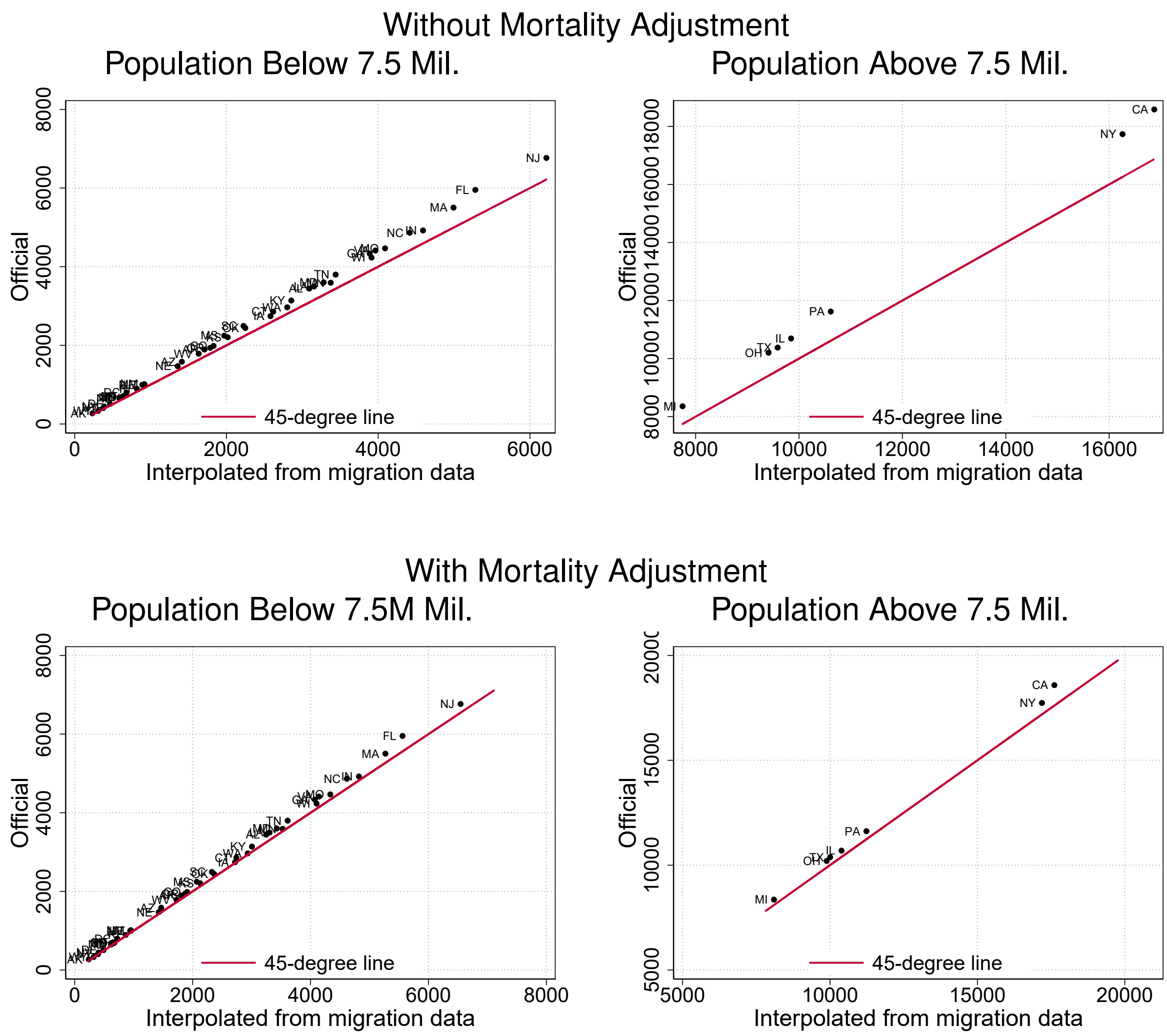

Figure A.1. Comparison of Official State-Level Population Levels to Aggregated Interpolated Estimates: 1965. *Note:* Each panel plots the total state-level populations from official US Census estimates on the vertical axis against total population imputed from migration data taken from the 1970 census (in which respondents were asked where they lived in 1965). The red line in each figure is the line along which the two populations would be equal. The left two panels show results for states with populations smaller than 7.5 million, and the right two panels show results for states with populations larger than 7.5 million. The top two panels show counts in which individuals who died during the 1965–1970 period are not counted in the imputed estimates. The bottom two panels show counts in which an estimate of the number of individuals who died during the 1965–1970 period is added to the imputed count. These estimates are derived by applying death rates by age, state, and gender to the population counts. *Source:* Authors' calculations using US Census Bureau data via IPUMS and mortality data from the United States Mortality Database.

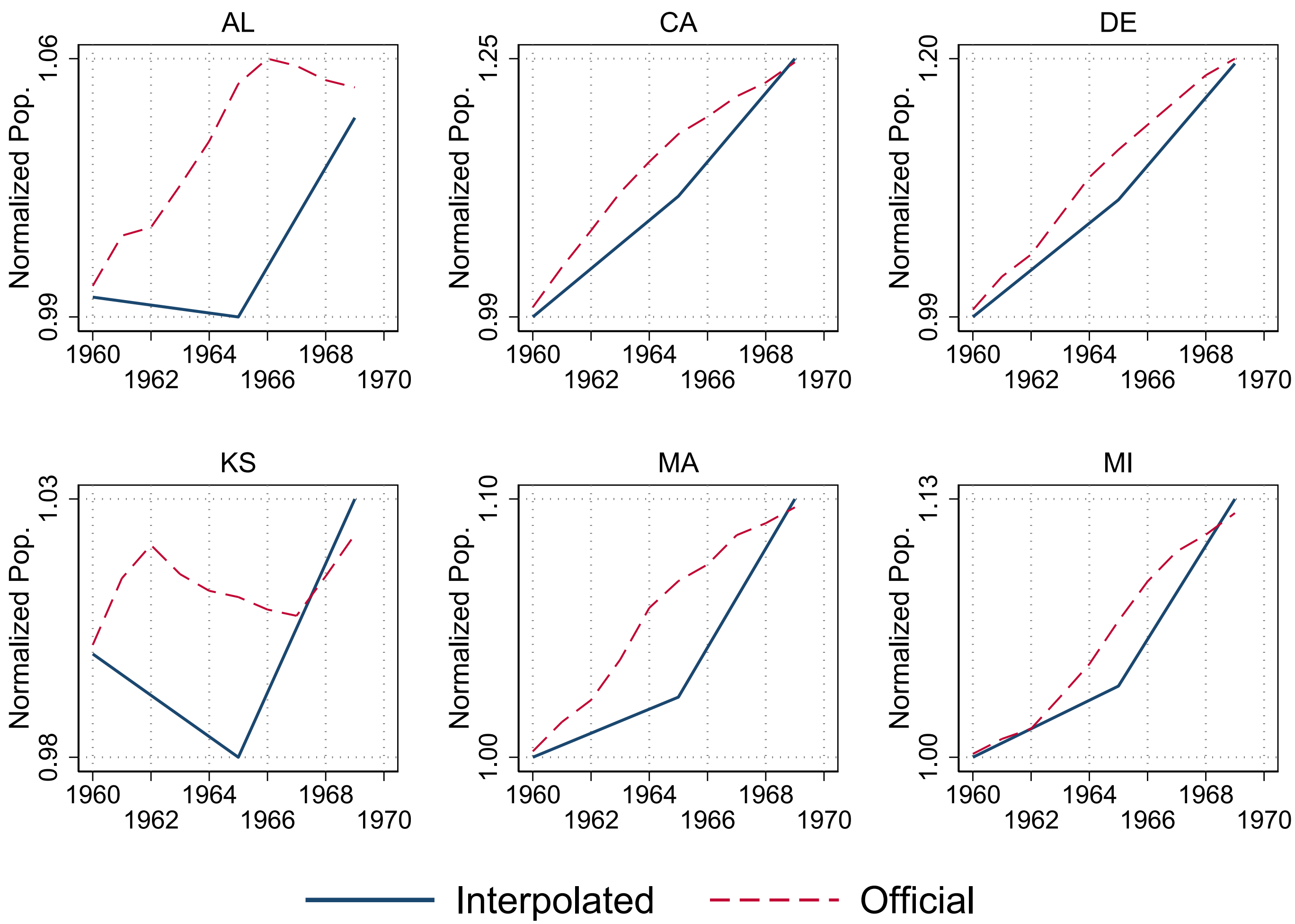

Figure A.2. Official Population Compared with Interpolated for Select States: 1960 through 1969. *Note:* Each panel shows an estimate of a state's population from official US Census estimates (the red dashed line) and an estimate constructed by the authors based on a linear interpolation between 1960, 1965, and 1969 (the solid blue line). In the interpolated data, the 1960 population estimate is based on US Census microdata from the 1960 census. The 1965 population is based on migration data from the 1970 census and an adjustment for mortality. The 1969 population is based on data from SEER. All estimates have been normalized such that the 1960 population is equal to one. *Source:* Authors' calculations using US Census Bureau data via IPUMS, mortality data from the United States Mortality Database and population data from SEER.

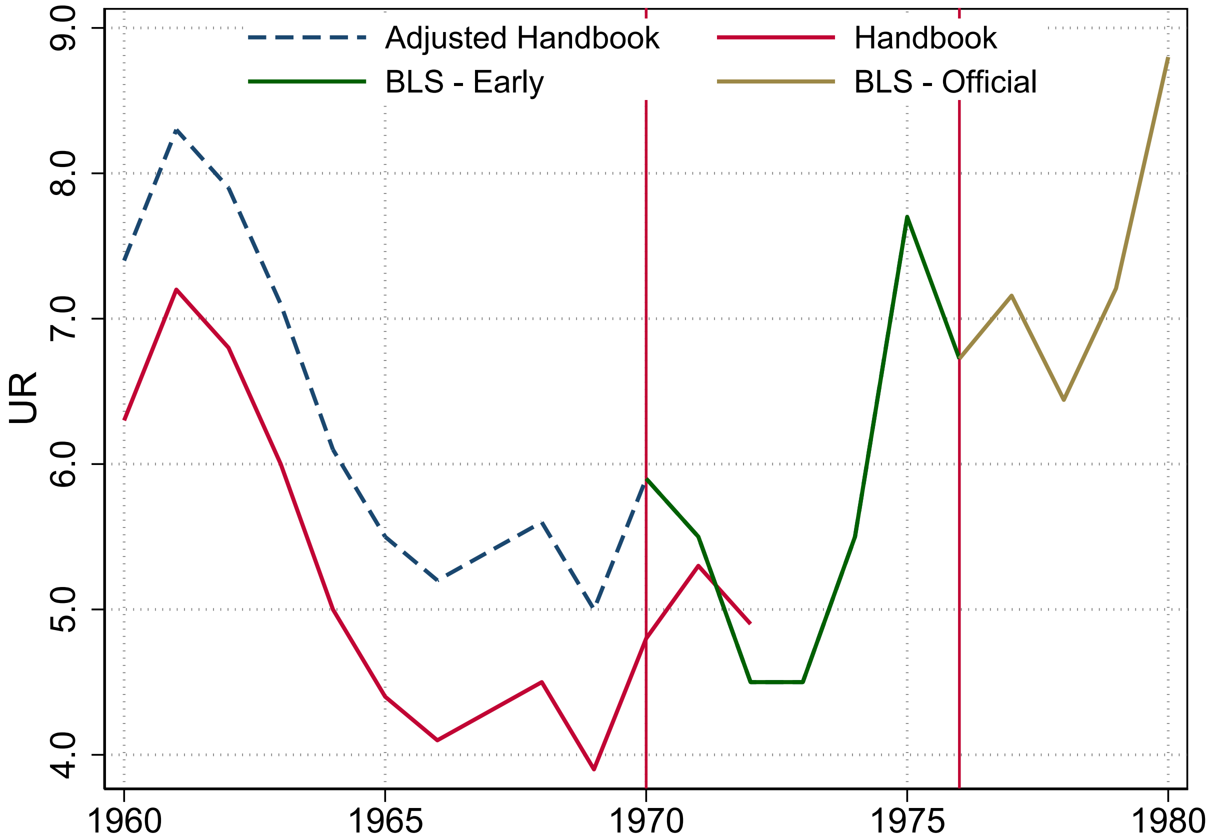

Figure A.3. Demonstration of UR Adjustment for Alabama. *Note:* The red line shows an estimate of the unemployment rate (UR) in Alabama from 1960 through 1972 that combines data taken from the *Manpower Reports of the President*. The green line shows estimates based on early, unofficial estimates calculated by the BLS and supplied by Susan Gorel to Robert Shimer. The light brown line shows official BLS estimates, which begin in 1976. The dashed blue line demonstrates a level adjustment applied to the *Manpower Report* series. The difference between the Shimer-Gorel estimate and the *Manpower Report* estimate is calculated in 1970 (the first year the Shimer-Gorel estimates for Alabama are available), then the difference is added to the entire *Manpower Report* series. *Source:* Authors' calculations using BLS data.

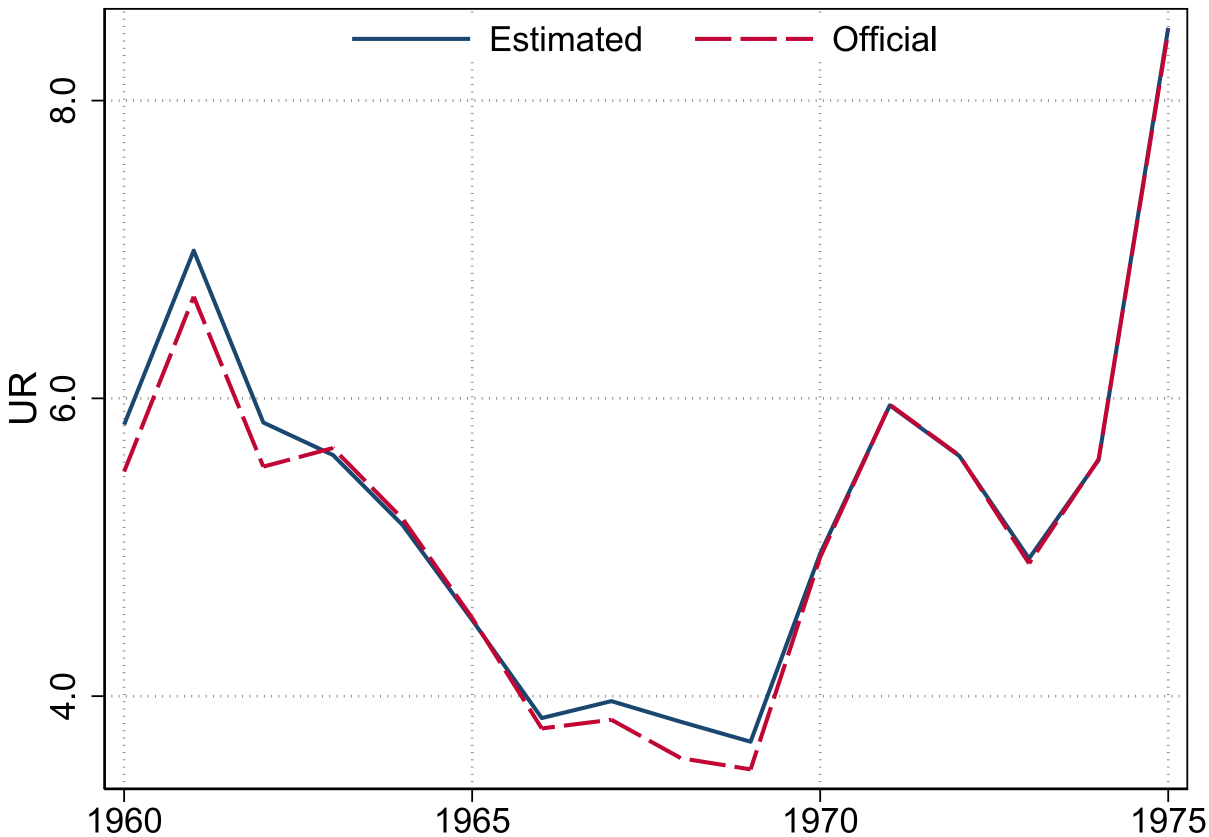

Figure A.4. Comparison of Aggregated National URs. *Note:* The blue line shows an estimate of the US unemployment rate (UR) based on a population-weighted average of state unemployment rates. The red dashed line shows the official national unemployment rate. *Source:* Authors' calculations using BLS data (unemployment rates) and SEER data (population).

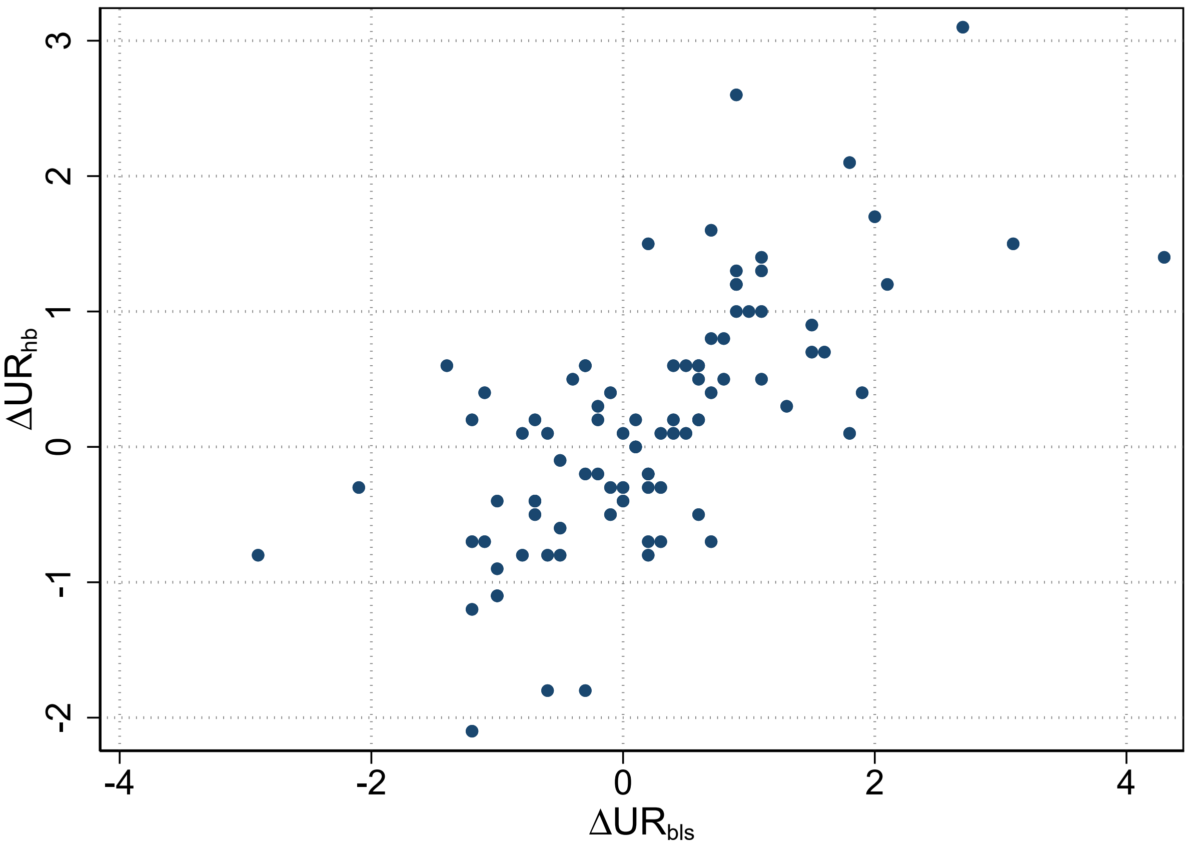

Figure A.5. Comparing Year-to-Year Differences in State-Level Unemployment Rates by Data Source. *Note:* This figure plots the year-to-year difference in the "handbook" unemployment rates taken from the *Manpower Reports* (vertical axis) against the same differences taken from the unofficial BLS estimates from the early 1970s (horizontal axis). The unofficial BLS estimates were supplied by Susan Gorel to Robert Shimer. Only state-year observations in which both differences can be calculated are displayed. *Source:* Authors' calculations using BLS data.

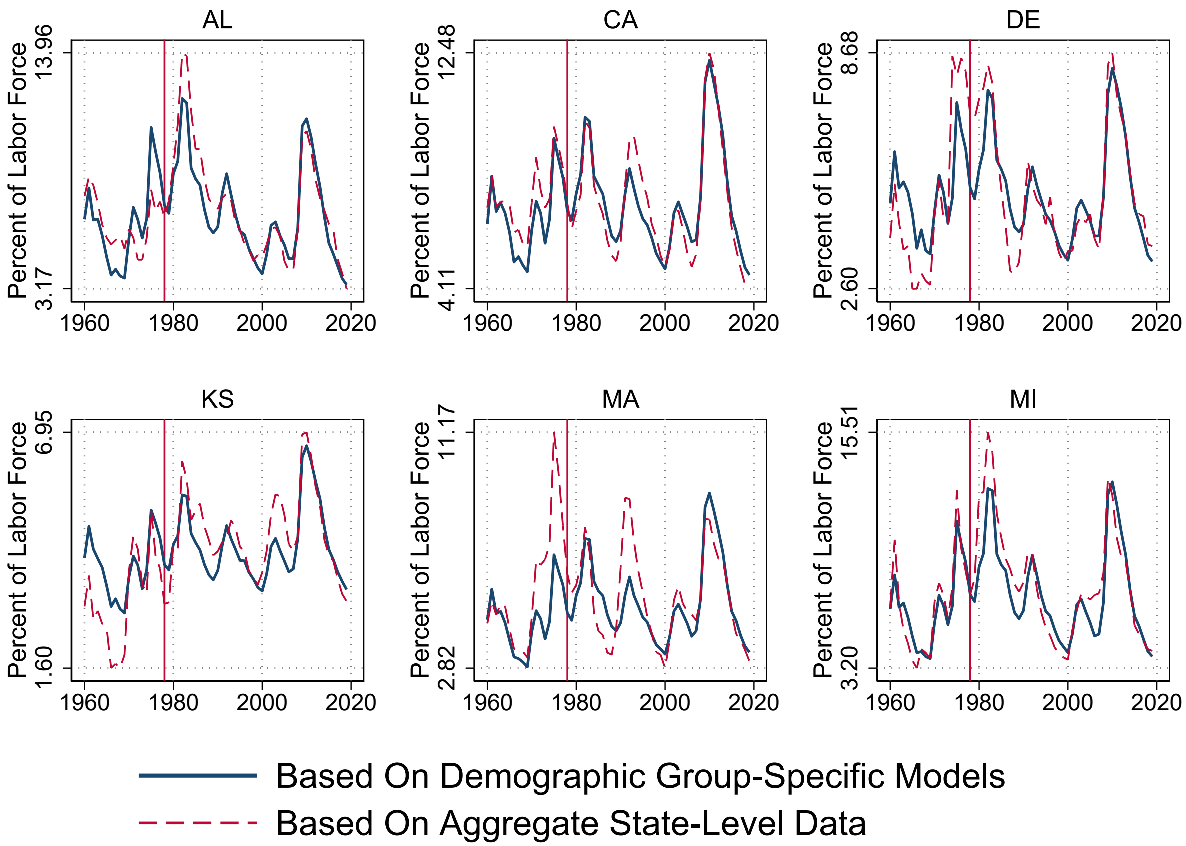

Figure A.6. Comparing Unemployment Rates Based on Group-Specific Models and Aggregate Data for a Selection of States. *Note:* Each panel plots two versions of a state's unemployment rate. The blue line shows a rate that is built up from predicted values generated by unemployment regressions run separately for demographic groups based on age, gender, and state. The left-hand-side variable in these models is the unemployment rate for the group taken from CPS microdata. The right-hand-side variables are the group-specific national unemployment rate and the overall national unemployment rate. The blue line is calculated by aggregating up the predicted values from these models using the group-specific population shares within each state. The vertical red line is drawn at 1978, the first year that the data needed for the group-specific regressions are available. The dashed red line shows a state-level unemployment rate that is based on aggregate data. This rate uses official BLS estimates of state-level unemployment after 1976 and our splicing calculation using the *Manpower Report* data and the Shimer-Gorel data before 1976. *Source:* Authors' calculations using BLS data and CPS microdata via IPUMS.