A New Claims-Based Unemployment Dataset: Application to Postwar Recoveries Across U.S. States

Andrew Fieldhouse
Mays Business School
Texas A&M University

Sean Howard
Wood Mackenzie

Christoffer Koch
International Monetary Fund

David Munro
Middlebury College

January 6, 2023

ASSA 2023 Annual Meeting

Motivation

Macroeconomists are increasingly leveraging panel datasets and regional heterogeneity to identify economic relationships

- Nakamura and Steinsson (2014); Chodorow-Reich (2019); Hazell, Nakamura and Steinsson (2022); Glandon et al (2022)

Motivation

Macroeconomists are increasingly leveraging panel datasets and regional heterogeneity to identify economic relationships

- Nakamura and Steinsson (2014); Chodorow-Reich (2019); Hazell, Nakamura and Steinsson (2022); Glandon et al (2022)

Recent papers have also emphasized that business cycle analysis should focus on the unemployment rate

- Romer and Romer (2019); Hall and Kudlyak (2020)

Motivation

Macroeconomists are increasingly leveraging panel datasets and regional heterogeneity to identify economic relationships

- Nakamura and Steinsson (2014); Chodorow-Reich (2019); Hazell, Nakamura and Steinsson (2022); Glandon et al (2022)

Recent papers have also emphasized that business cycle analysis should focus on the unemployment rate

- Romer and Romer (2019); Hall and Kudlyak (2020)

Regrettably, official state-level unemployment data only begin in 1976, a significant impediment to historical state-level analyses

Contributions

We digitize monthly state-level unemployment claims data back to 1947 from a series of primary sources

Contributions

We digitize monthly state-level unemployment claims data back to 1947 from a series of primary sources

Using this data, we construct claims-based unemployment rates, which are highly correlated with official measures

- Monthly data Jan 1947-May 2022, for all 50 states, DC, US
- Nearly three additional decades of monthly state-level data

Contributions

We digitize monthly state-level unemployment claims data back to 1947 from a series of primary sources

Using this data, we construct claims-based unemployment rates, which are highly correlated with official measures

- Monthly data Jan 1947-May 2022, for all 50 states, DC, US
- Nearly three additional decades of monthly state-level data

With this new dataset we explore various features of post-war U.S. recessions at the national and state level

- Backdated data span the first six post-war U.S. recessions
- Faster national labor market recoveries in the 1940s, 50s were associated with greater dispersion of recovery rates across states
- States with larger manufacturing sectors tend to see faster recoveries

Data Digitization and Construction

Digitization Overview

We digitize monthly state-level data on Initial Claims (IC) and Continued Claims (CC) from various government reports:

- *Employment Security Activities*, *The Labor Market and Employment Security*, *Unemployment Insurance Statistics*, *Unemployment Insurance Review*

Digitization Overview

We digitize monthly state-level data on Initial Claims (IC) and Continued Claims (CC) from various government reports:

- *Employment Security Activities*, *The Labor Market and Employment Security*, *Unemployment Insurance Statistics*, *Unemployment Insurance Review*

We tracked down these primary sources/scanned tables from HathiTrust, Google Books, Interlibrary Loan requests, and scans from the Department of Labor's internal library

Digitization Overview

We digitize monthly state-level data on Initial Claims (IC) and Continued Claims (CC) from various government reports:

- *Employment Security Activities*, *The Labor Market and Employment Security*, *Unemployment Insurance Statistics*, *Unemployment Insurance Review*

We tracked down these primary sources/scanned tables from HathiTrust, Google Books, Interlibrary Loan requests, and scans from the Department of Labor's internal library

Digitized data on monthly unemployment claims is available online from the Department of Labor's website for 1971+

Digitization Overview

We digitize monthly state-level data on Initial Claims (IC) and Continued Claims (CC) from various government reports:

- *Employment Security Activities*, *The Labor Market and Employment Security*, *Unemployment Insurance Statistics*, *Unemployment Insurance Review*

We tracked down these primary sources/scanned tables from HathiTrust, Google Books, Interlibrary Loan requests, and scans from the Department of Labor's internal library

Digitized data on monthly unemployment claims is available online from the Department of Labor's website for 1971+

In total, we digitized just over 36,000 monthly observations

▸ Data Quality

Claims-based Unemployment Rates

Measuring Unemployment Rates

Unemployment rate is computed as: $UR = \frac{U}{E+U}$

Measuring Unemployment Rates

Unemployment rate is computed as: $UR = \frac{U}{E+U}$

Official national UR estimated from CPS ($\sim$ 60k households)

Measuring Unemployment Rates

Unemployment rate is computed as: $UR = \frac{U}{E+U}$

Official national UR estimated from CPS ($\sim$ 60k households)

At the state level, official UR (1976+) is not a measured object, but a statistical construct

- Constructed from CPS, unemployment claims, other employment surveys, statistical filtering

Measuring Unemployment Rates

Unemployment rate is computed as: $UR = \frac{U}{E+U}$

Official national UR estimated from CPS ($\sim$ 60k households)

At the state level, official UR (1976+) is not a measured object, but a statistical construct

- Constructed from CPS, unemployment claims, other employment surveys, statistical filtering

Our claims data is an alternative – conceptually similar yet different – measure of U

- Use initial, continued claims as a measure of unemployed workers

Claims-Based Unemployment Rates

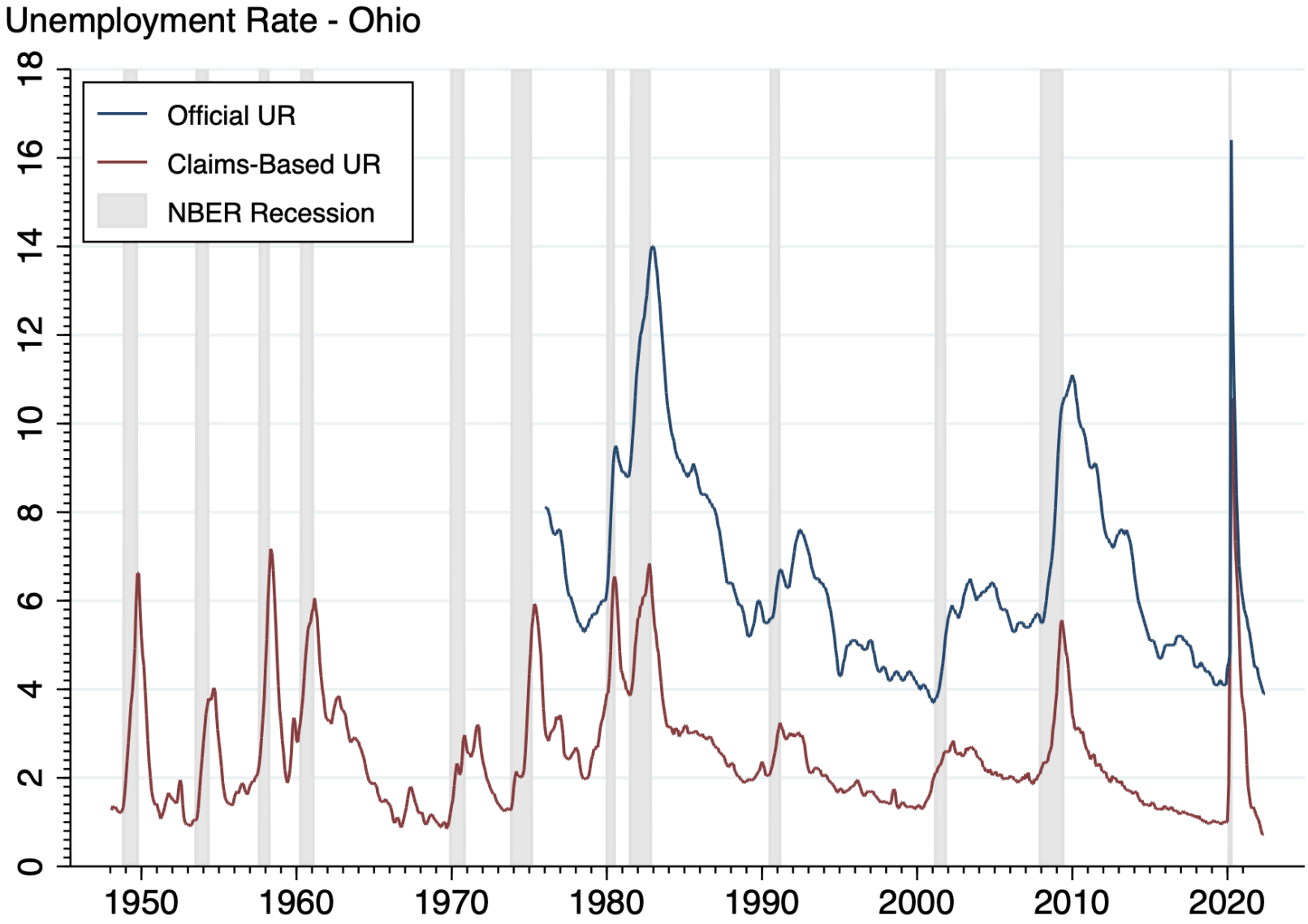
Our claims-based unemployment rate for state i in month t is computed as

$$UR_{i,t}^{Claims} = \frac{IC_{i,t} + CC_{i,t}}{NP_{i,t} + IC_{i,t} + CC_{i,t}} \tag{1}$$

- Where are $IC + CC$ is our proxy for U
- We use nonfarm payroll (NP) employment as our measure of E (only measure of state-level employment to 1940s)

▸ Data Frequency

Claims-Based Unemployment Rate Example: Ohio



Claims-Based Unemployment Rate: Ohio Unpacked

Correlation for overlapping sample: 0.82

Claims-Based Unemployment Rate: Ohio Unpacked

Correlation for overlapping sample: 0.82

Practical benefit:

- Our data series provides roughly three decades of additional data
- Data spans six additional national recessions (1948-49 – 1973-75)

Claims-Based Unemployment Rate: Ohio Unpacked

Correlation for overlapping sample: 0.82

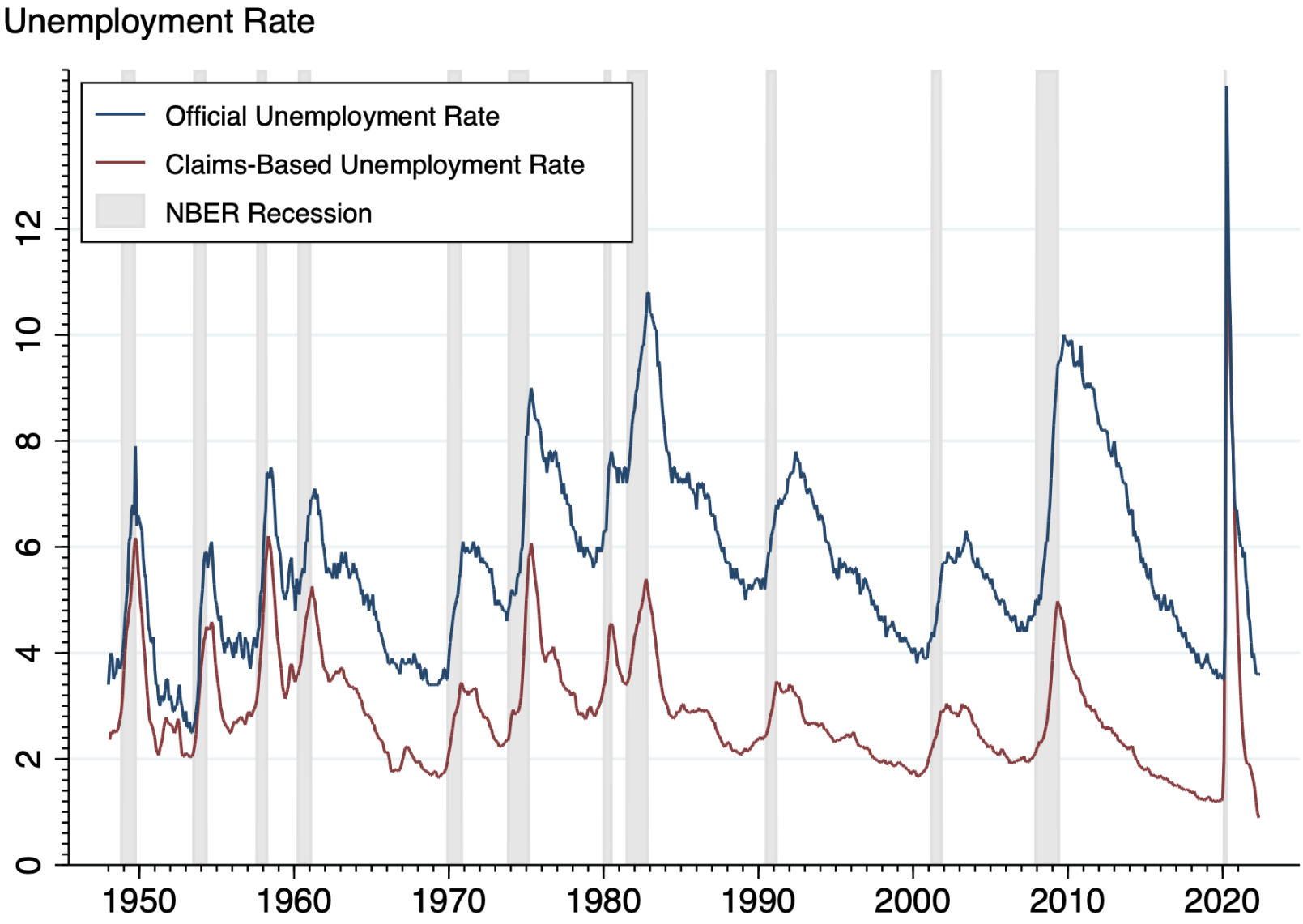
Practical benefit:

- Our data series provides roughly three decades of additional data
- Data spans six additional national recessions (1948-49 – 1973-75)

Level difference to be expected:

- Narrower pool of benefit-eligible workers, benefit exhaustion
- Shouldn't matter for business cycle analysis so long as series are highly correlated, identify similar inflection points

Claims-Based Unemployment Rates: National



▸ Total Employment

State Business Cycles

Business Cycle Properties of the Data

Using this data we investigate various features of post-war recessions at both national and state level

Business Cycle Properties of the Data

Using this data we investigate various features of post-war recessions at both national and state level

Lots of recent attention paid to the pace of economic recoveries, esp. unemployment recoveries

- Dupraz, Nakamura, Steinsson (2019), Hall and Kudlyak (2022)

Business Cycle Properties of the Data

Using this data we investigate various features of post-war recessions at both national and state level

Lots of recent attention paid to the pace of economic recoveries, esp. unemployment recoveries

- Dupraz, Nakamura, Steinsson (2019), Hall and Kudlyak (2022)

Using our data we first examine the timing and pace of national recoveries as litmus test

- Our claims-based unemployment rate picks up consistent business cycle features as BLS national unemployment rate

Business Cycle Properties of the Data

Using this data we investigate various features of post-war recessions at both national and state level

Lots of recent attention paid to the pace of economic recoveries, esp. unemployment recoveries

- Dupraz, Nakamura, Steinsson (2019), Hall and Kudlyak (2022)

Using our data we first examine the timing and pace of national recoveries as litmus test

- Our claims-based unemployment rate picks up consistent business cycle features as BLS national unemployment rate

After we have some confidence in our claims-based unemployment rates, we explore state-level recoveries

Recovery Rates and Recession Dating

Following Hall and Kudlyak (2022) we compute the pace of recovery as mean decline in log unemployment over recovery:

$$\text{Recovery Pace} = -12 \cdot (\log UR_0 - \log UR_T)/T$$

Recovery Rates and Recession Dating

Following Hall and Kudlyak (2022) we compute the pace of recovery as mean decline in log unemployment over recovery:

$$\text{Recovery Pace} = -12 \cdot (\log UR_0 - \log UR_T)/T$$

We first need to date business cycles first to determine start of recovery (0) and end of recovery (T) for each expansion

Recovery Rates and Recession Dating

Following Hall and Kudlyak (2022) we compute the pace of recovery as mean decline in log unemployment over recovery:

$$\text{Recovery Pace} = -12 \cdot (\log UR_0 - \log UR_T)/T$$

We first need to date business cycles first to determine start of recovery (0) and end of recovery (T) for each expansion

We adopt the relatively simple, unemployment-based recession dating algorithm proposed in Dupraz, Nakamura, and Steinsson (2019) (DNS, henceforth) ▸ DNS

- Generates a close match to NBER dates, Hall and Kudlyak (2022) chronology of unemployment-based recession dates

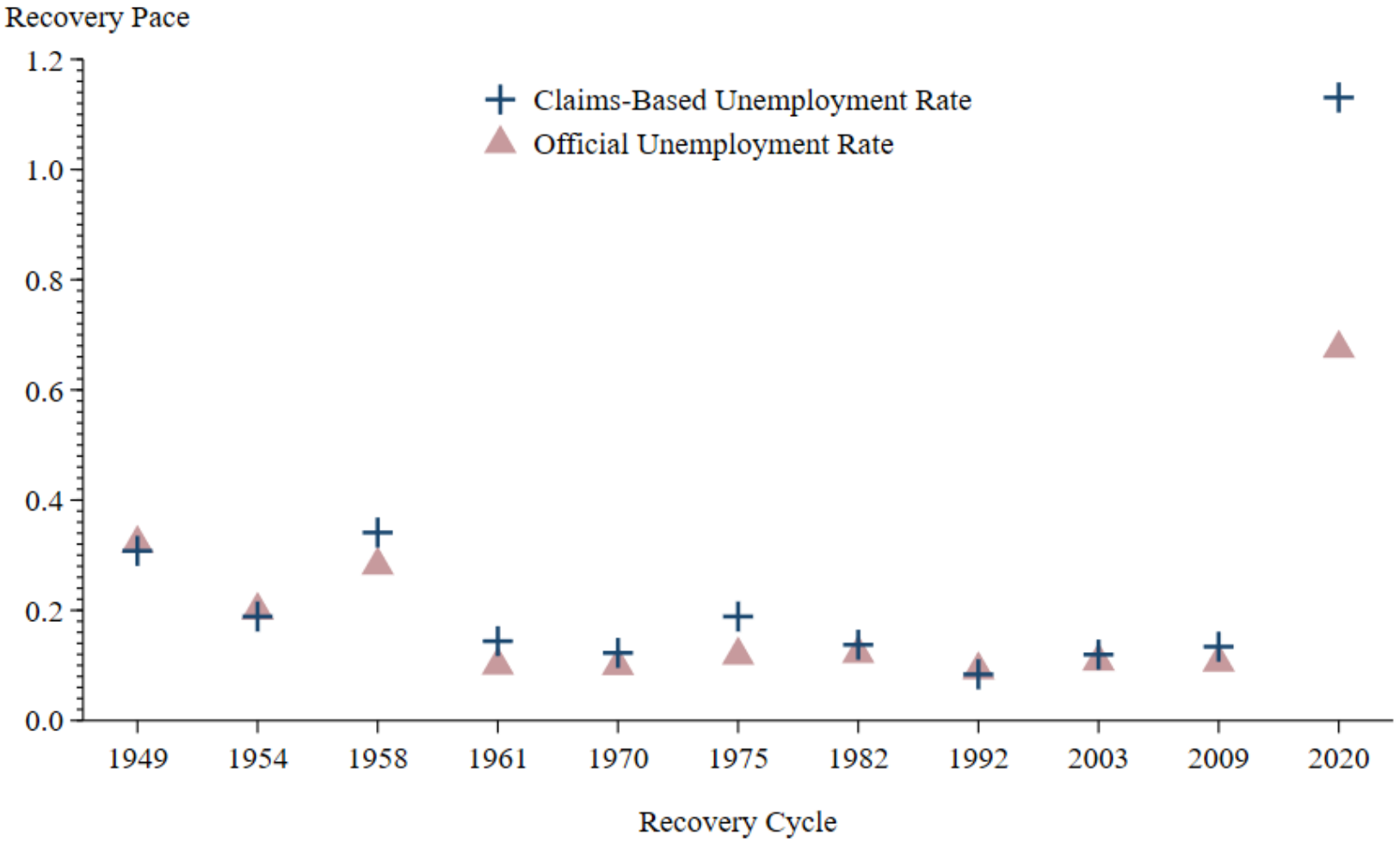
Recession Dating: National

Table 1: Business Cycle Peaks and Troughs

			DNS Dating Algorithm			
	NBER		Claims-based UR		Official UR	
	Peak	Trough	Peak	Trough	Peak	Trough
1	Nov. 1948	Oct. 1949	[Feb. 1948]	Oct. 1949	[Jan. 1948]	Oct. 1949
2	July 1953	May 1954	Apr. 1953	Sep. 1954	May 1953	Sep. 1954
3	Aug. 1957	Apr. 1958	Dec. 1955	May 1958	Mar. 1957	July 1958
4	Apr. 1960	Feb. 1961	June 1959	Mar. 1961	Feb. 1960	May 1961
5	Dec. 1969	Nov. 1970	June 1969	Nov. 1970	Sep. 1968	Dec. 1970
6	Nov. 1973	Mar. 1975	Apr. 1973	May 1975	Oct. 1973	May 1975
7a	Jan. 1980	July 1980	Nov. 1978	July 1980	May 1979	
7b	July 1981	Nov. 1982	June 1981	Oct. 1982		Nov. 1982
8	July 1990	Mar. 1991	Nov. 1988	Mar. 1991	Mar. 1989	June 1992
9	Mar. 2001	Nov. 2001	Apr. 2000	Mar. 2002	Apr. 2000	June 2003
10	Dec. 2007	June 2009	Apr. 2006	May 2009	Oct. 2006	Oct. 2009
11	[Feb. 2020]	Apr. 2020	June 2019	May 2020	Sep. 2019	Apr. 2020

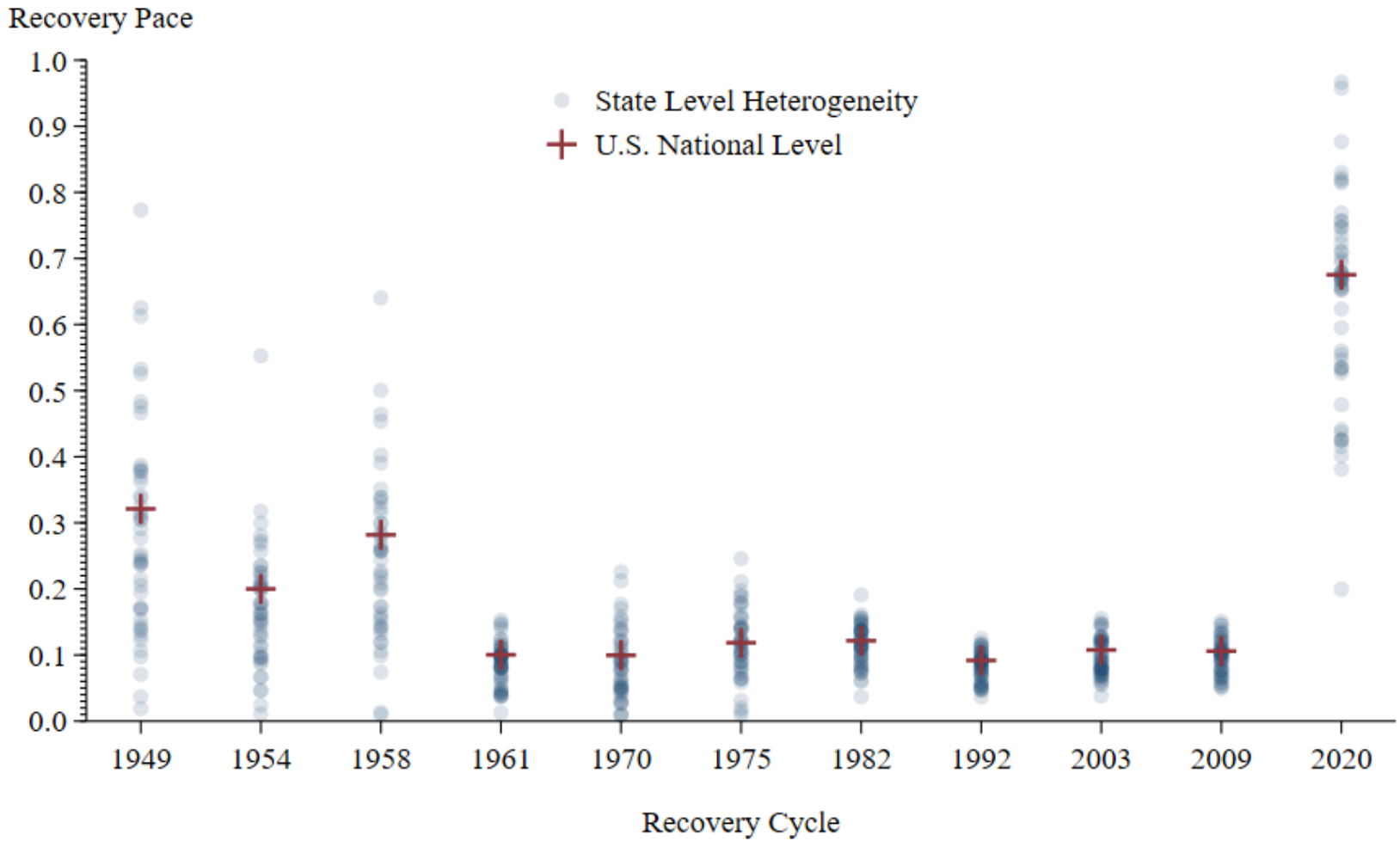
Notes: Recession dates for CBUR and UR are generated by applying the DNS algorithm on these two series. For the UR, we use the DNS parameter of 1.5. For CBUR we choose a parameter of 1.0, which is able to capture the NBER recession events.

Recovery Pace: National Recoveries

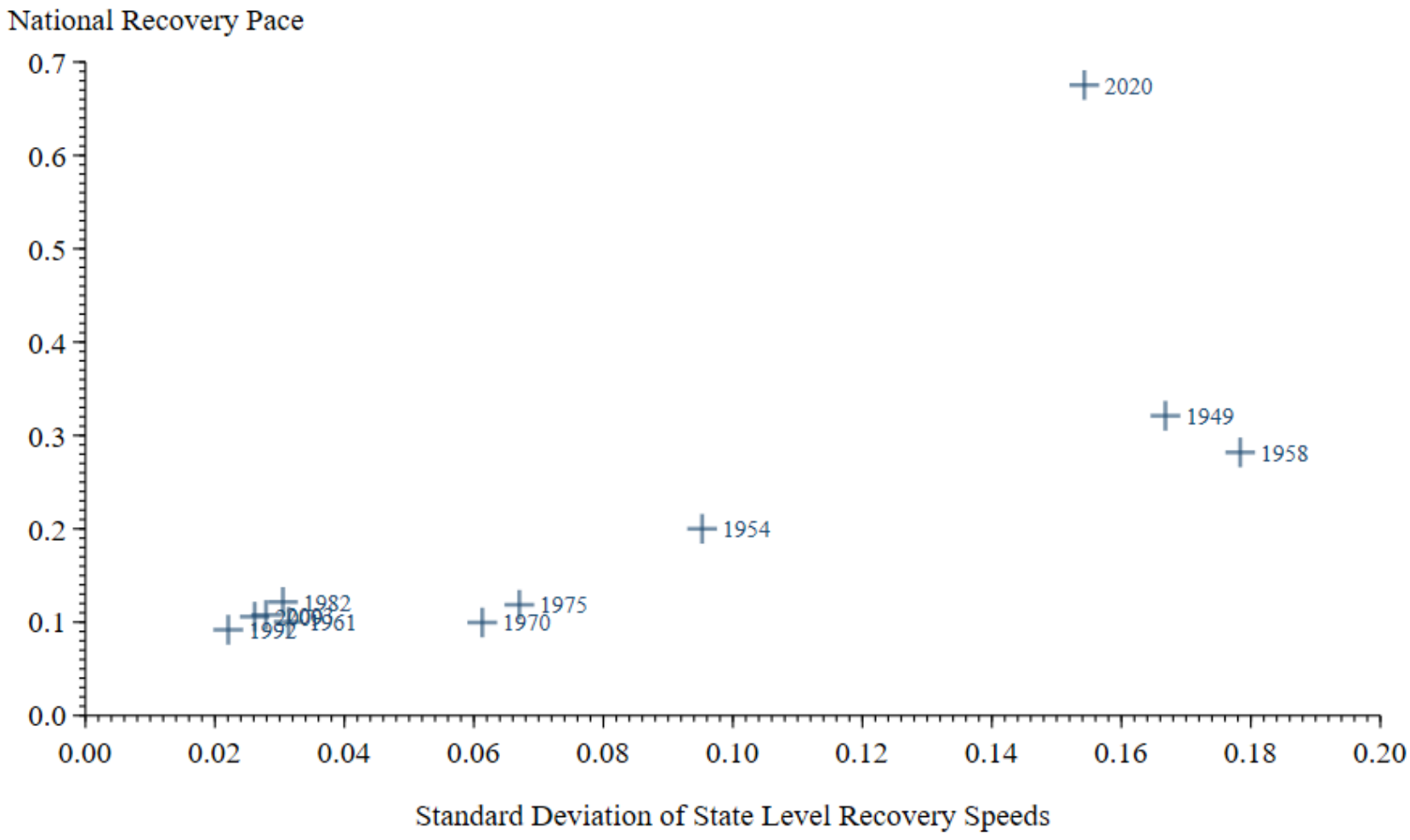


▸ CBUR Recession Dates

Recovery Pace: State-level Recoveries



Recovery Pace: National Rate vs. State-level Dispersion



State Recovery Rate Takeaways

Recession dates and the pace of recoveries at the national level using our claims-based unemployment rates line-up quite well with analogous results using the official unemployment rate

State Recovery Rate Takeaways

Recession dates and the pace of recoveries at the national level using our claims-based unemployment rates line-up quite well with analogous results using the official unemployment rate

State-level analysis reveals some interesting heterogeneity

State Recovery Rate Takeaways

Recession dates and the pace of recoveries at the national level using our claims-based unemployment rates line-up quite well with analogous results using the official unemployment rate

State-level analysis reveals some interesting heterogeneity

Of course, with state-level data you can begin to think about what other factors correlate with features of the business cycle

State Recovery Rate Takeaways

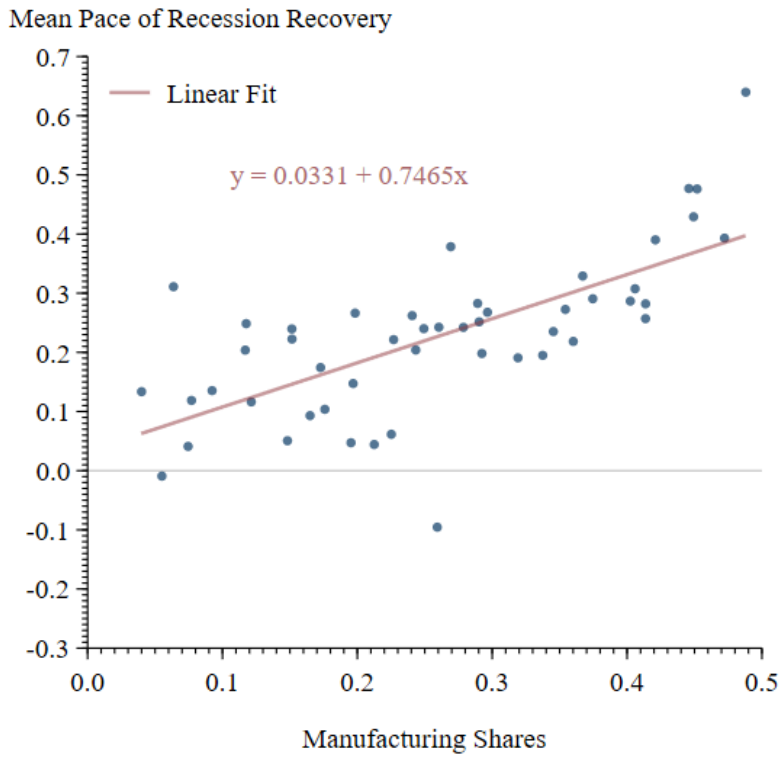
Recession dates and the pace of recoveries at the national level using our claims-based unemployment rates line-up quite well with analogous results using the official unemployment rate

State-level analysis reveals some interesting heterogeneity

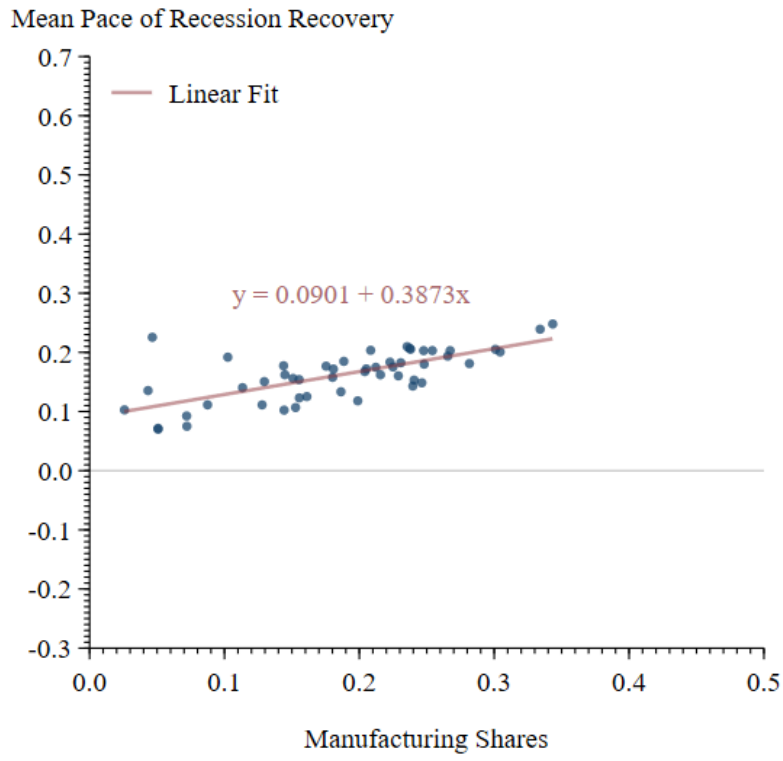
Of course, with state-level data you can begin to think about what other factors correlate with features of the business cycle

One thing that jumped out to us: the pace of recoveries is strongly correlated with the size of states' manufacturing sector

Recovery Pace by State Manufacturing Share



1949, '54, '58 Recoveries



1961-2009 Recoveries

Concluding Thoughts

Conclusion

- We digitize state-level unemployment claims data back to 1947 to expand our historical record of unemployment

Conclusion

- We digitize state-level unemployment claims data back to 1947 to expand our historical record of unemployment
- With this historical claims data we construct claims-based unemployment rates: highly correlated w/ official unemployment rates, similar business cycle features

Conclusion

- We digitize state-level unemployment claims data back to 1947 to expand our historical record of unemployment
- With this historical claims data we construct claims-based unemployment rates: highly correlated w/ official unemployment rates, similar business cycle features
- Our claims-based unemployment series doubles the number of post-war recessions that can be studied at state level

Conclusion

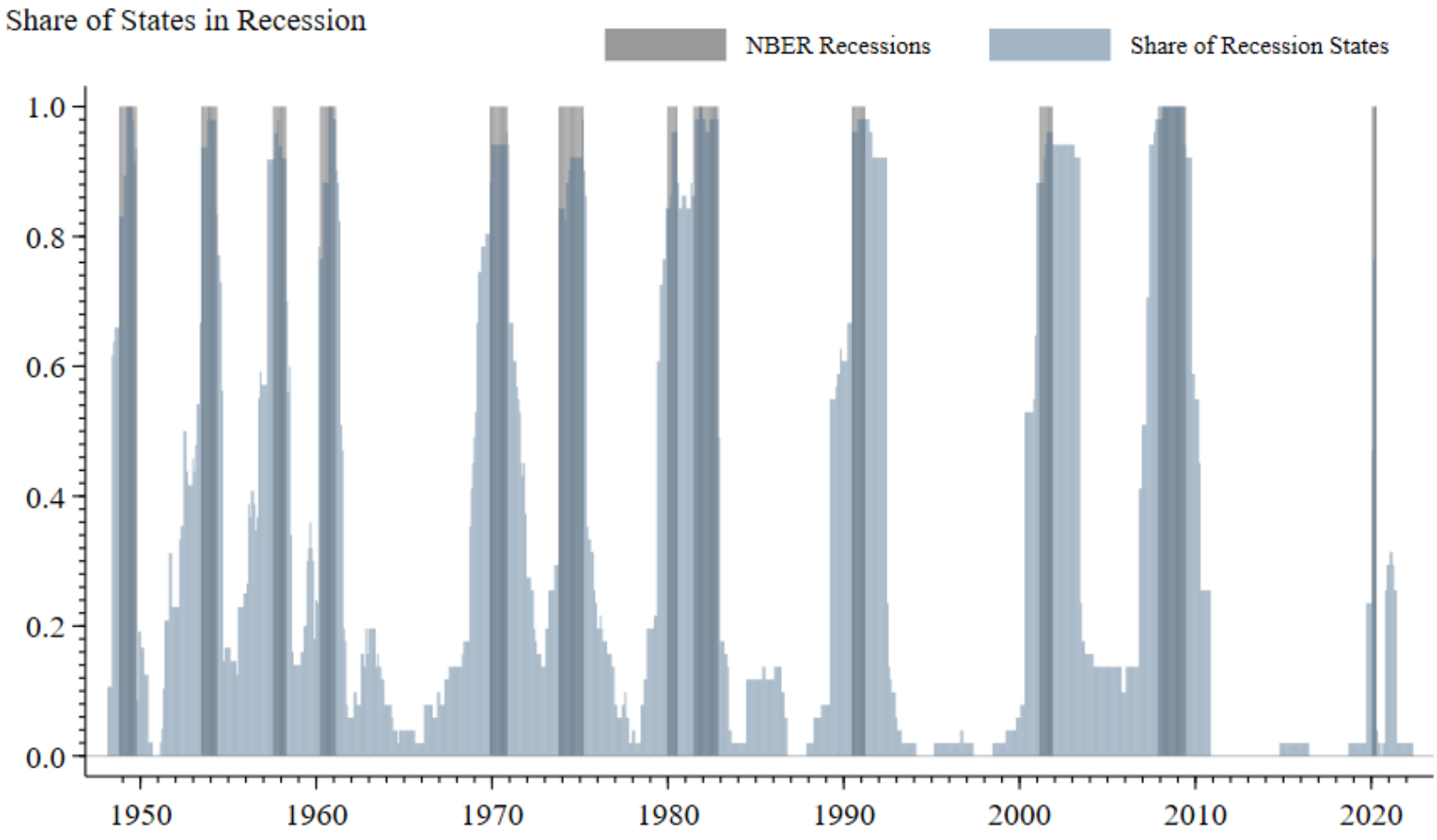
- We digitize state-level unemployment claims data back to 1947 to expand our historical record of unemployment
- With this historical claims data we construct claims-based unemployment rates: highly correlated w/ official unemployment rates, similar business cycle features
- Our claims-based unemployment series doubles the number of post-war recessions that can be studied at state level
- As a first pass, we use this data to study the timing and pace of post-war economic recoveries for U.S. states

Conclusion

- We digitize state-level unemployment claims data back to 1947 to expand our historical record of unemployment
- With this historical claims data we construct claims-based unemployment rates: highly correlated w/ official unemployment rates, similar business cycle features
- Our claims-based unemployment series doubles the number of post-war recessions that can be studied at state level
- As a first pass, we use this data to study the timing and pace of post-war economic recoveries for U.S. states
- The data could be used for a whole host of other questions, and we're excited about follow-up work

Appendix Slides

Recession Dating: State-level Recessions vs. NBER



▸ CBUR Recession Dates

Digitization and Data Quality

Overall, legibility of the scans we tracked down was quite good

Digitization and Data Quality

Overall, legibility of the scans we tracked down was quite good

When we encountered legibility issues we

- Obtained secondary images
- Used national aggregates to detect inconsistencies
- Used percentage change from previous month or annual changes to detect errors
- All digitizations had multiple sets of eyes review them

Digitization and Data Quality

Overall, legibility of the scans we tracked down was quite good

When we encountered legibility issues we

- Obtained secondary images
- Used national aggregates to detect inconsistencies
- Used percentage change from previous month or annual changes to detect errors
- All digitizations had multiple sets of eyes review them

Very few sum-check errors, digitization merged quite well

Digitization and Data Quality

Overall, legibility of the scans we tracked down was quite good

When we encountered legibility issues we

- Obtained secondary images
- Used national aggregates to detect inconsistencies
- Used percentage change from previous month or annual changes to detect errors
- All digitizations had multiple sets of eyes review them

Very few sum-check errors, digitization merged quite well

We seasonally adjust the digitized, existing data (Win X-13)

▸ Outliers ▸ Back

Outliers

We seasonally adjust the digitized data (Census Win X-13)

This process also identified ∼200 outliers from roughly 91,000 observations, roughly evenly distributed between our newly digitized data and the existing DOL data

▸ Back

Outliers

We seasonally adjust the digitized data (Census Win X-13)

This process also identified ~200 outliers from roughly 91,000 observations, roughly evenly distributed between our newly digitized data and the existing DOL data

Each outlier was manually checked to evaluate if it was a legitimate change in claims or a "fat thumb" coding error

- Example of legitimate outlier: surge in LA post-Katrina
- Example of "fat thumb" error: In MO June 1974 CC surged 4700% from 147,351 to 7,132,843 then back to 145,365: Population of MO was less than 5 million

Back

Outliers

We seasonally adjust the digitized data (Census Win X-13)

This process also identified ~200 outliers from roughly 91,000 observations, roughly evenly distributed between our newly digitized data and the existing DOL data

Each outlier was manually checked to evaluate if it was a legitimate change in claims or a "fat thumb" coding error

- Example of legitimate outlier: surge in LA post-Katrina
- Example of "fat thumb" error: In MO June 1974 CC surged 4700% from 147,351 to 7,132,843 then back to 145,365: Population of MO was less than 5 million

We used our best judgement in fixing the "fat thumb" errors

▸ Back

Claims-Based Unemployment Rates: Data Frequency

The digitized monthly IC, CC data reflect all claims filed with the state unemployment office in that month

▸ Back

Claims-Based Unemployment Rates: Data Frequency

The digitized monthly IC, CC data reflect all claims filed with the state unemployment office in that month

Double counting concern: An individual can show up as both an initial claimant and a continued claimant and/or show up repeatedly as a continued claimant in a given month

▸ Back

Claims-Based Unemployment Rates: Data Frequency

The digitized monthly IC, CC data reflect all claims filed with the state unemployment office in that month

Double counting concern: An individual can show up as both an initial claimant and a continued claimant and/or show up repeatedly as a continued claimant in a given month

To avoid such double-counting of individuals, we convert monthly claims to average weekly claims for $IC_{i,t}$, $CC_{i,t}$

▸ Back

Claims-Based Unemployment Rates: Data Frequency

The digitized monthly IC, CC data reflect all claims filed with the state unemployment office in that month

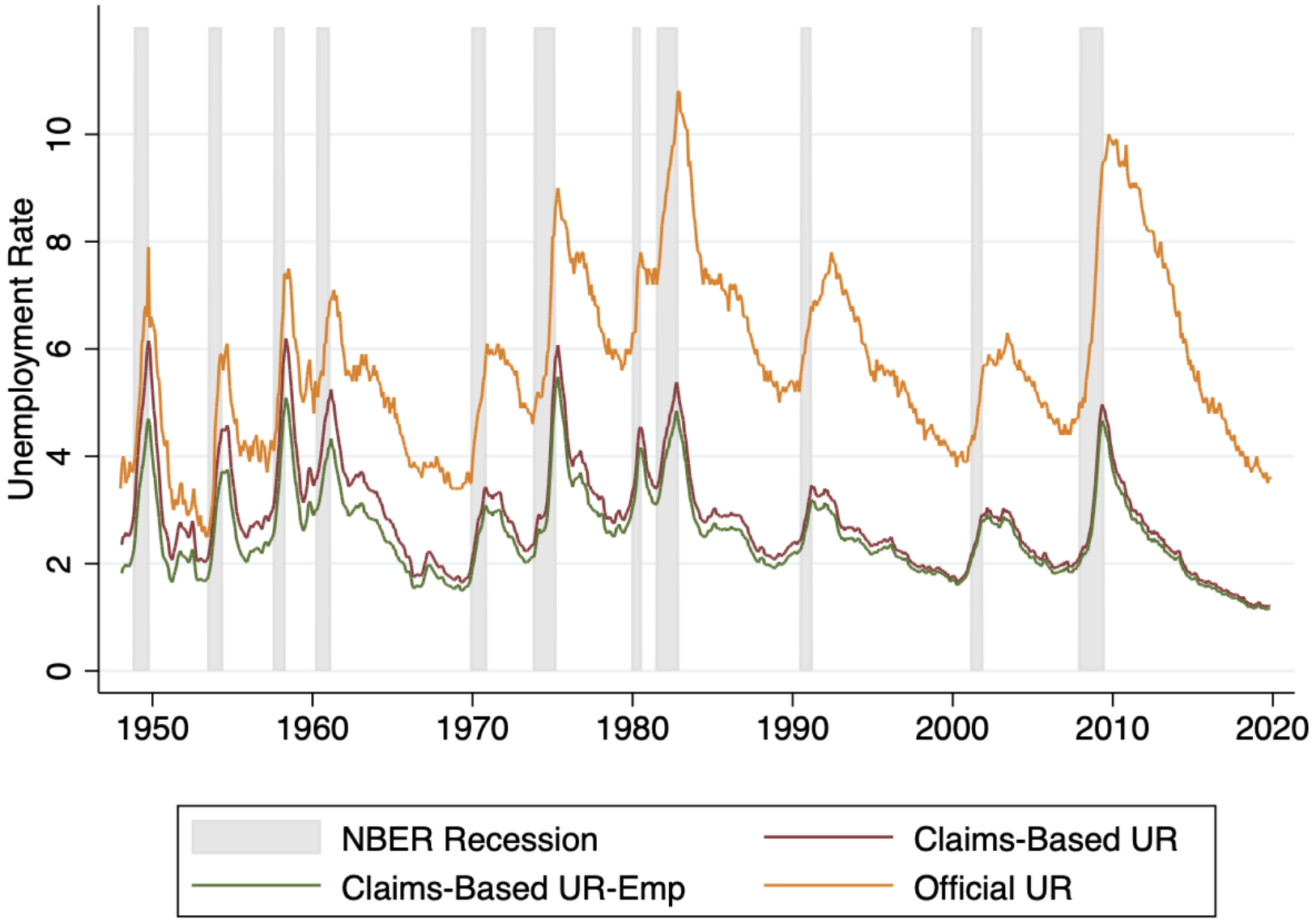
Double counting concern: An individual can show up as both an initial claimant and a continued claimant and/or show up repeatedly as a continued claimant in a given month

To avoid such double-counting of individuals, we convert monthly claims to average weekly claims for $IC_{i,t}$, $CC_{i,t}$

- Conceptually approach similar to the BLS's reference week used in sampling labor force activity, DOL's insured unemployment
- Monthly data are weighted by the split number of five-day workweeks in the month (weights as the sum or workdays in each given month, ignoring holidays, divided by five)

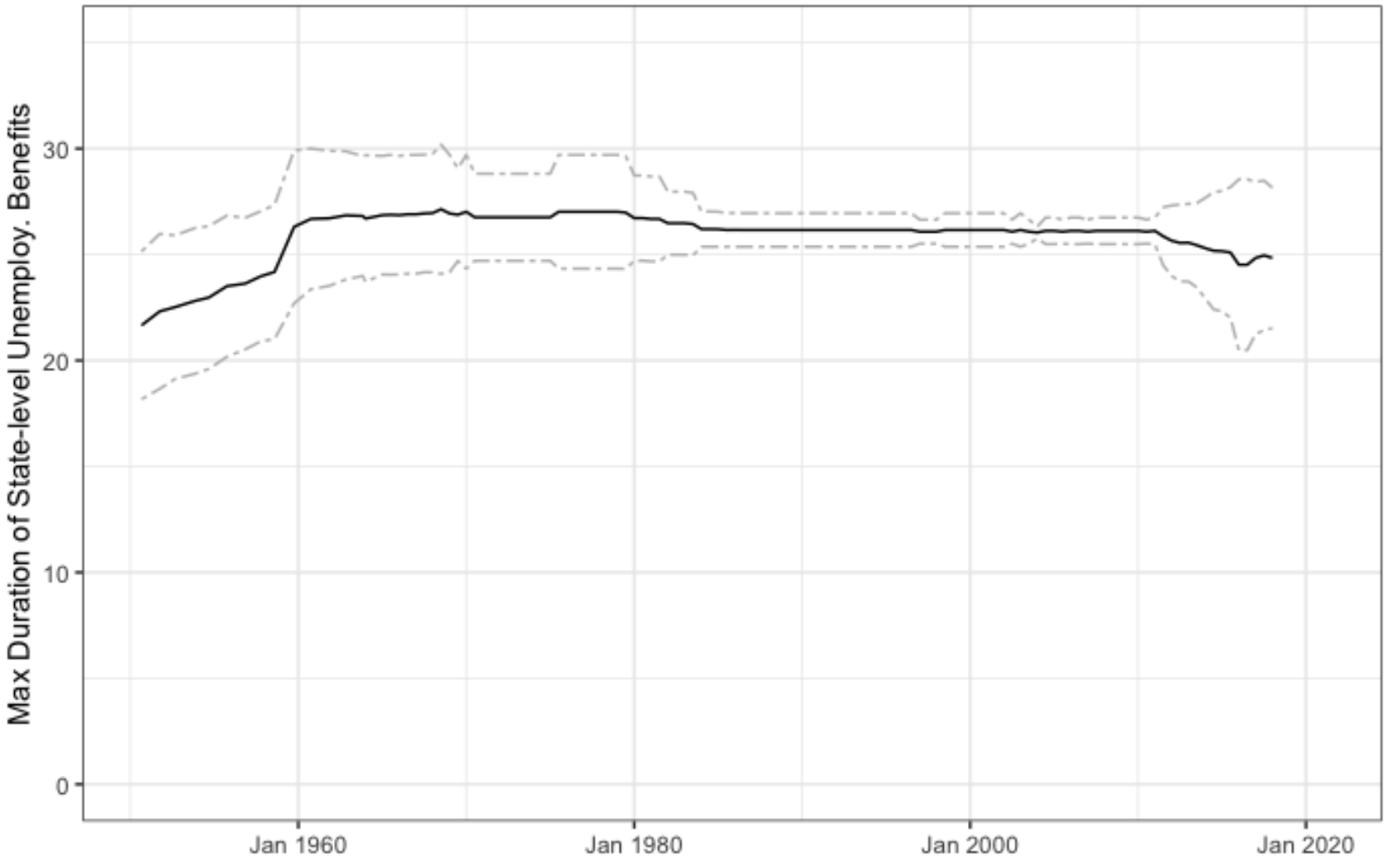
▸ Back

Claims-Based Unemployment Rates: Total Employment



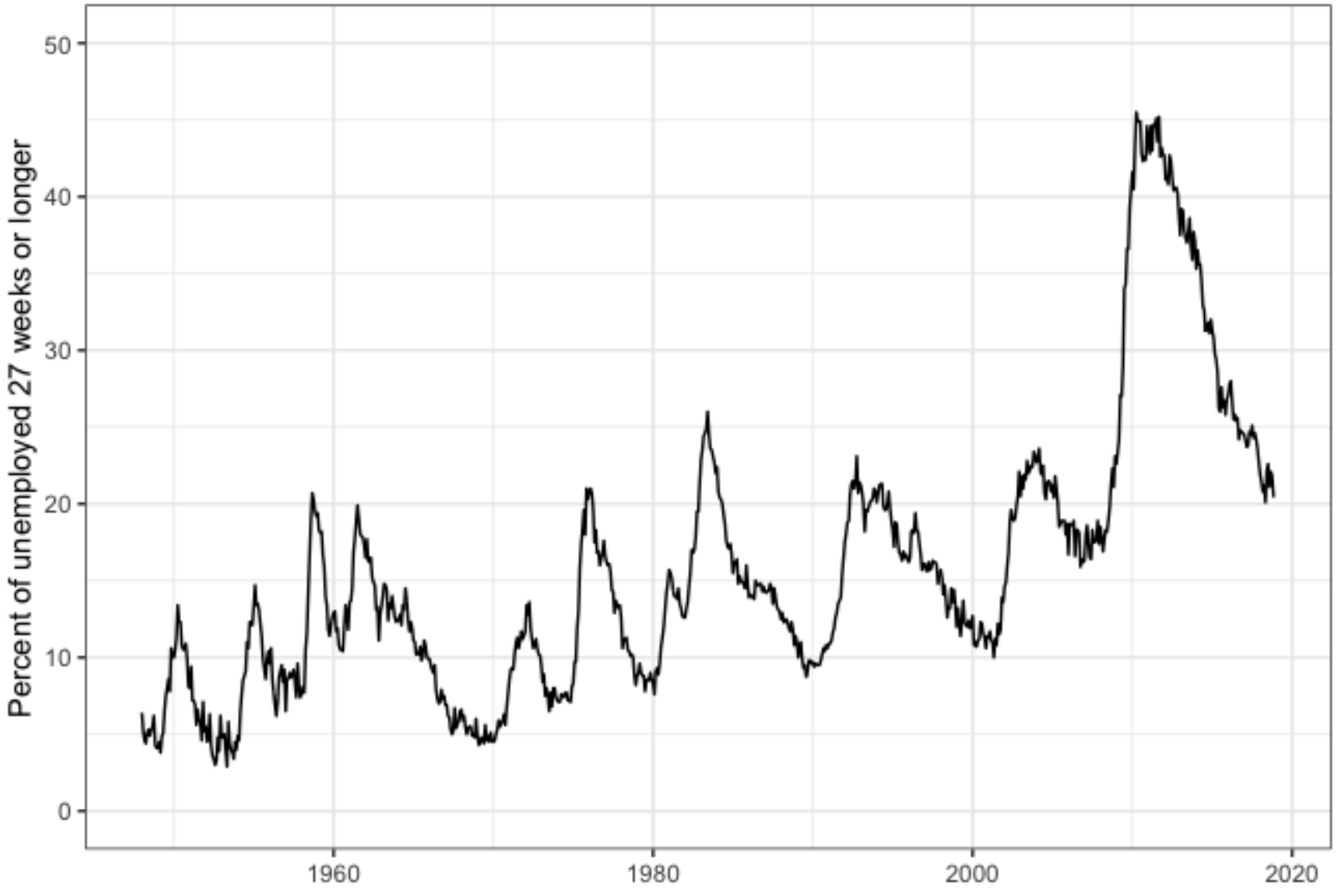
▸ Back

State-level Max Duration



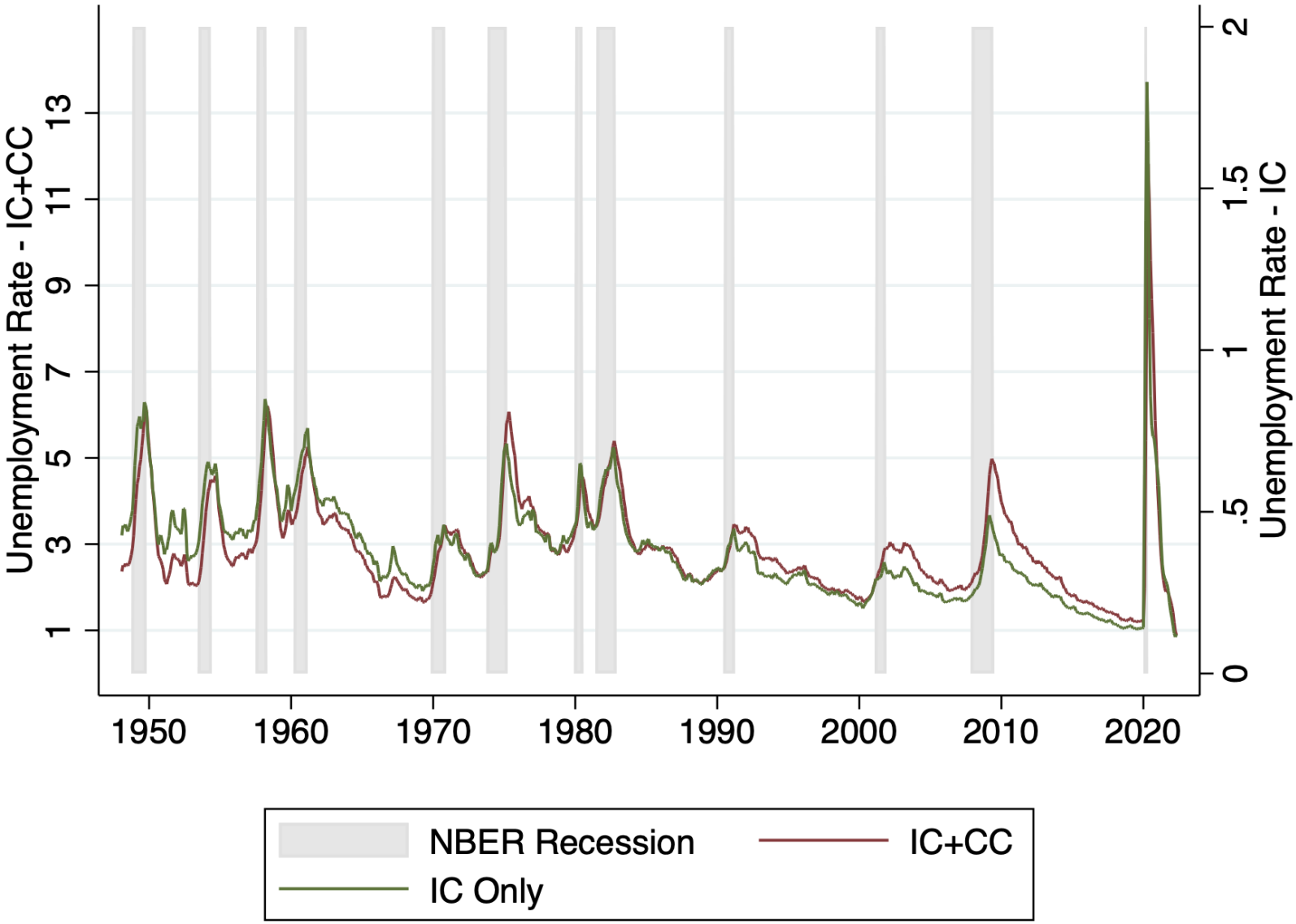
Back

Long-Term Unemployment Share



▸ Back

Alt. Claims-Based Unemployment Rate: IC Only



Back

Comparison with the Insured Unemployment Rate

Our claims-based unemployment rates lie conceptually between BLS's UR and DOL's Insured unemployment rate (IUR)

$$IUR = \frac{\text{Average Weekly } CC}{\text{Lagged Covered Employment}}$$

Comparison with the Insured Unemployment Rate

Our claims-based unemployment rates lie conceptually between BLS's UR and DOL's Insured unemployment rate (IUR)

$$IUR = \frac{\text{Average Weekly } CC}{\text{Lagged Covered Employment}}$$

- IUR also omits workers based on benefit eligibility, exhaustion, doesn't take a stance on search requirements

Comparison with the Insured Unemployment Rate

Our claims-based unemployment rates lie conceptually between BLS's UR and DOL's Insured unemployment rate (IUR)

$$IUR = \frac{\text{Average Weekly } CC}{\text{Lagged Covered Employment}}$$

- IUR also omits workers based on benefit eligibility, exhaustion, doesn't take a stance on search requirements
- IUR, $CBUR$ are highly correlated, close in levels

Comparison with the Insured Unemployment Rate

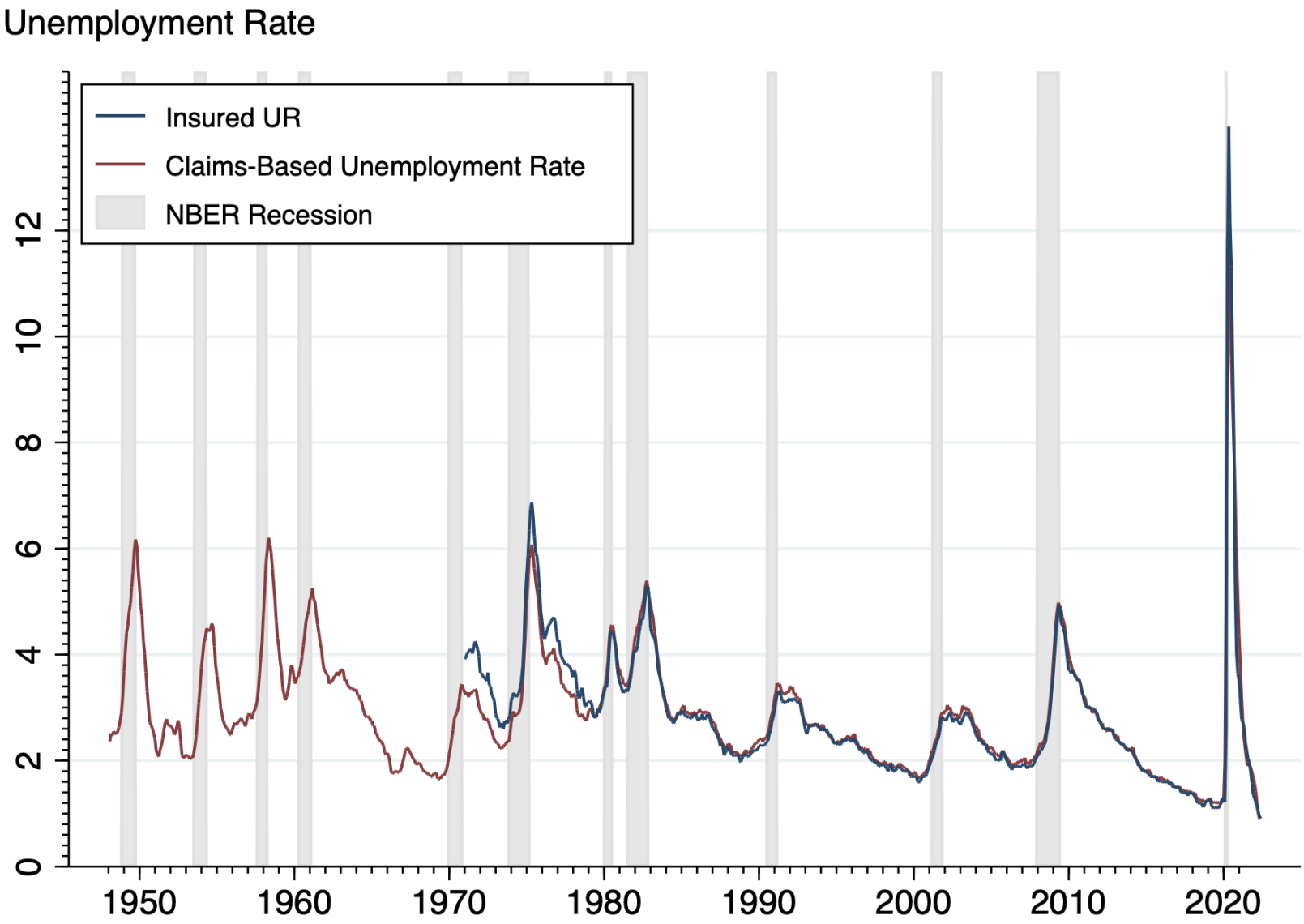
Our claims-based unemployment rates lie conceptually between BLS's UR and DOL's Insured unemployment rate (IUR)

$$IUR = \frac{\text{Average Weekly } CC}{\text{Lagged Covered Employment}}$$

- IUR also omits workers based on benefit eligibility, exhaustion, doesn't take a stance on search requirements
- IUR, $CBUR$ are highly correlated, close in levels
- But monthly IUR is only available for 1986+ at state level, 1971+ at national level

Back

U.S. Claims-Based, Insured Unemployment Rates



▸ Back

Fitted Model: Intuition and Performance

Fitting exercise captures simple intuition: a state's official unemployment rate is likely higher than national rate when they have a higher claims-based unemployment rate than national

Fitted Model: Intuition and Performance

Fitting exercise captures simple intuition: a state's official unemployment rate is likely higher than national rate when they have a higher claims-based unemployment rate than national

Also anchors estimates around the national unemployment rate, remove level differences, smooth series

Fitted Model: Intuition and Performance

Fitting exercise captures simple intuition: a state's official unemployment rate is likely higher than national rate when they have a higher claims-based unemployment rate than national

Also anchors estimates around the national unemployment rate, remove level differences, smooth series

These simple regressions fit official state-level URs very well:

- Avg. R^2=0.83
- Avg. correlation coefficient $= 0.91, \in (0.81 - 0.97)$

▸ Back

Recession Dating: DNS Algorithm

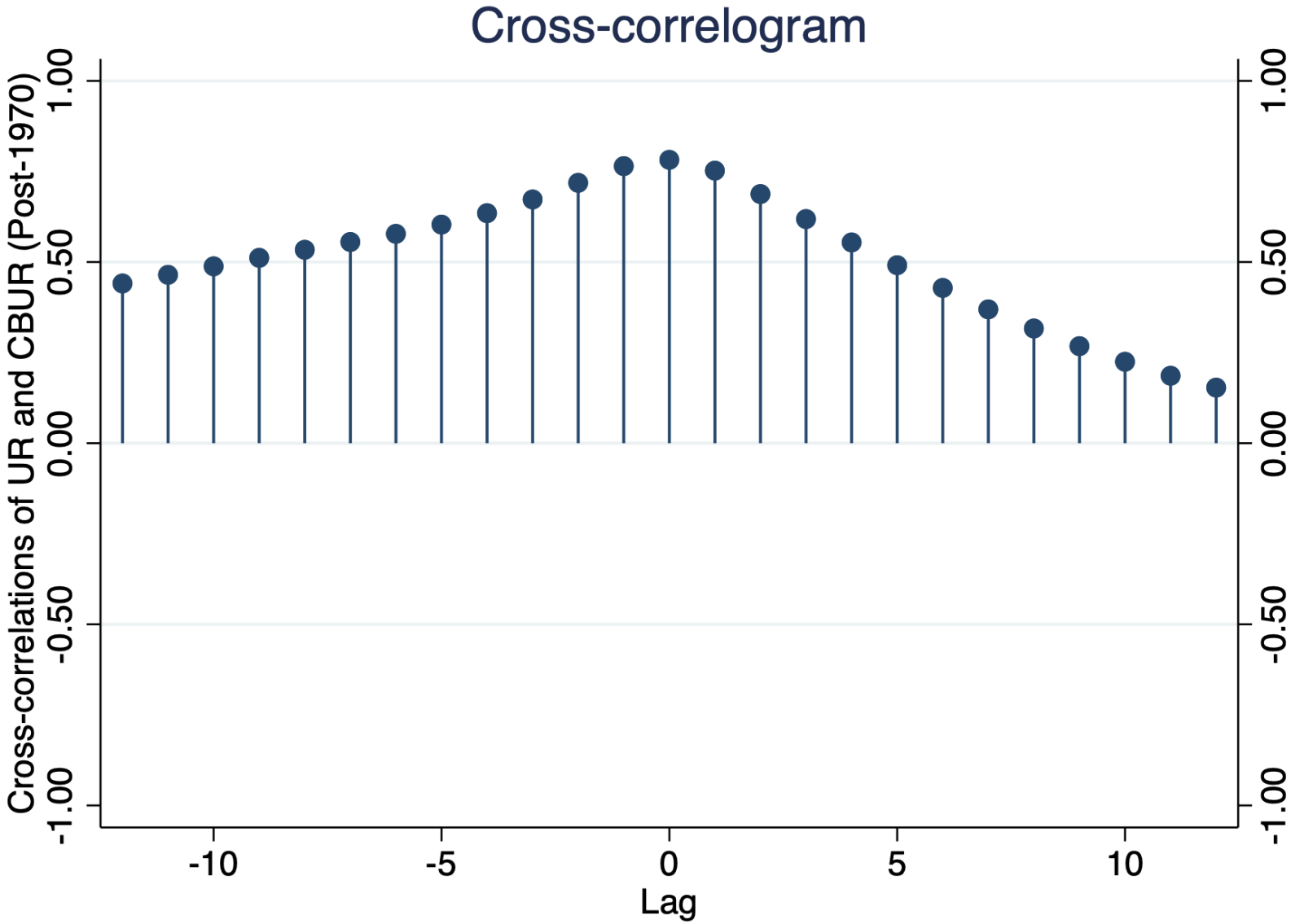
Gist: identifying local minima and maxima of the unemployment rate, ignoring low frequency variation in the unemployment rate

- Let u_t be a candidate for a cycle peak (cp)
- If $u_{t+h} > u_{cp}$ in all subsequent months until $u_{t+h+1} > u_{cp} + X$, confirm cp
- If $u_{t+h} < u_{cp}$, new candidate for cp
- After identifying a cp, proceed analogously to identify the next cycle trough (ct)...

Setting $X = 1.5$ identifies unemployment-based peak/troughs similar to those identified by NBER

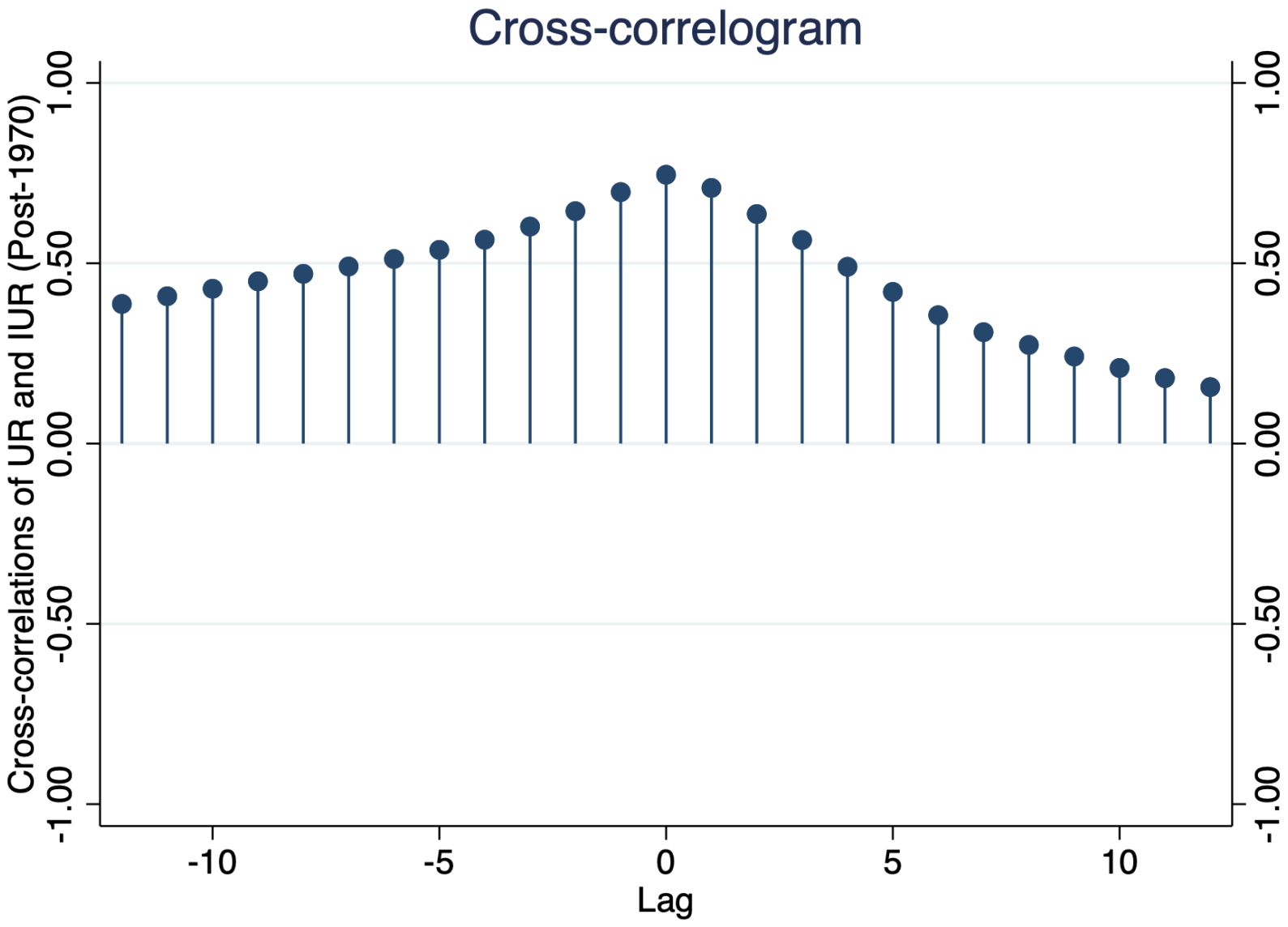
Back

Unemployment Rate-CBUR Cross Correlations



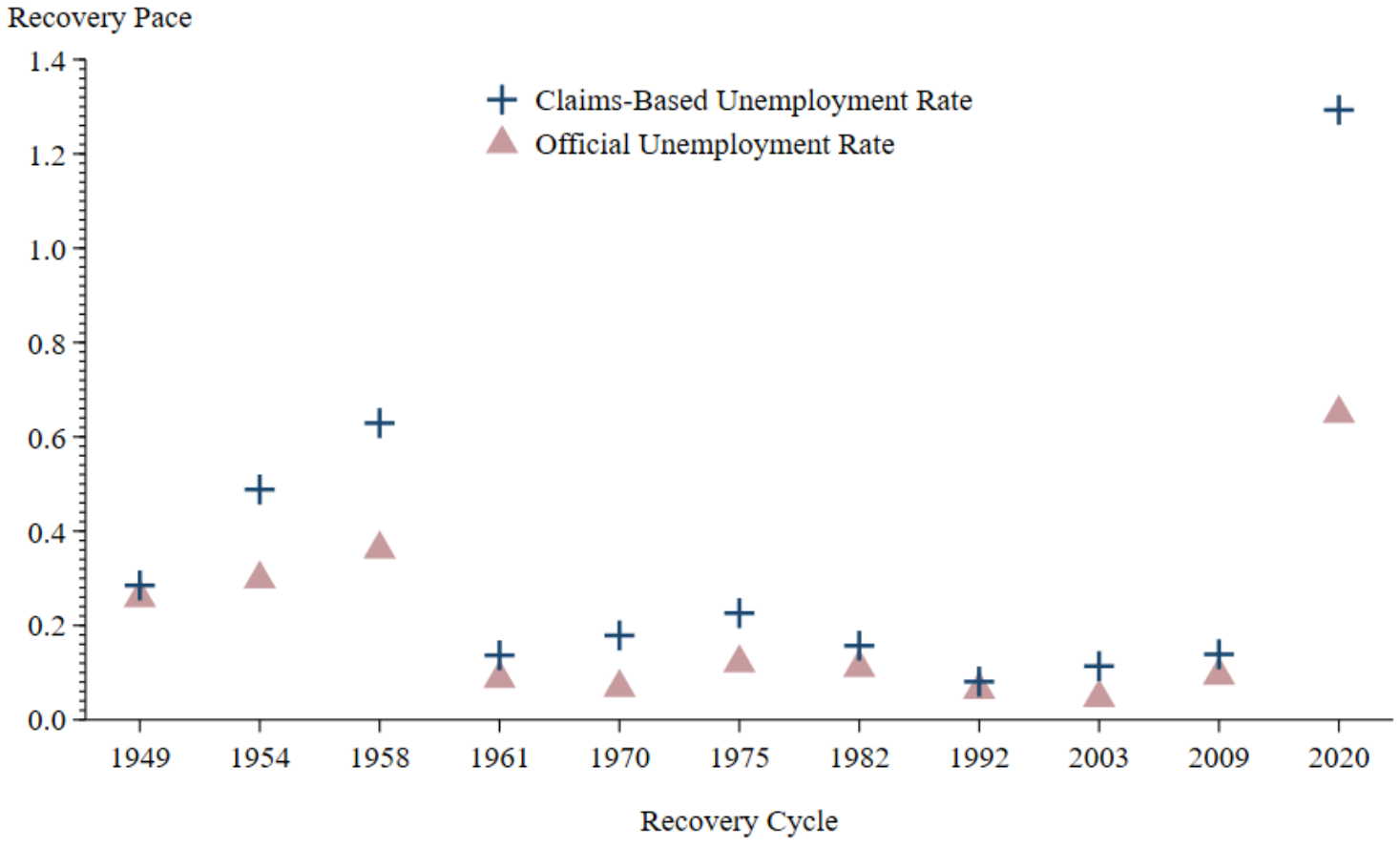
▸ Back

Unemployment Rate-IUR Cross Correlations



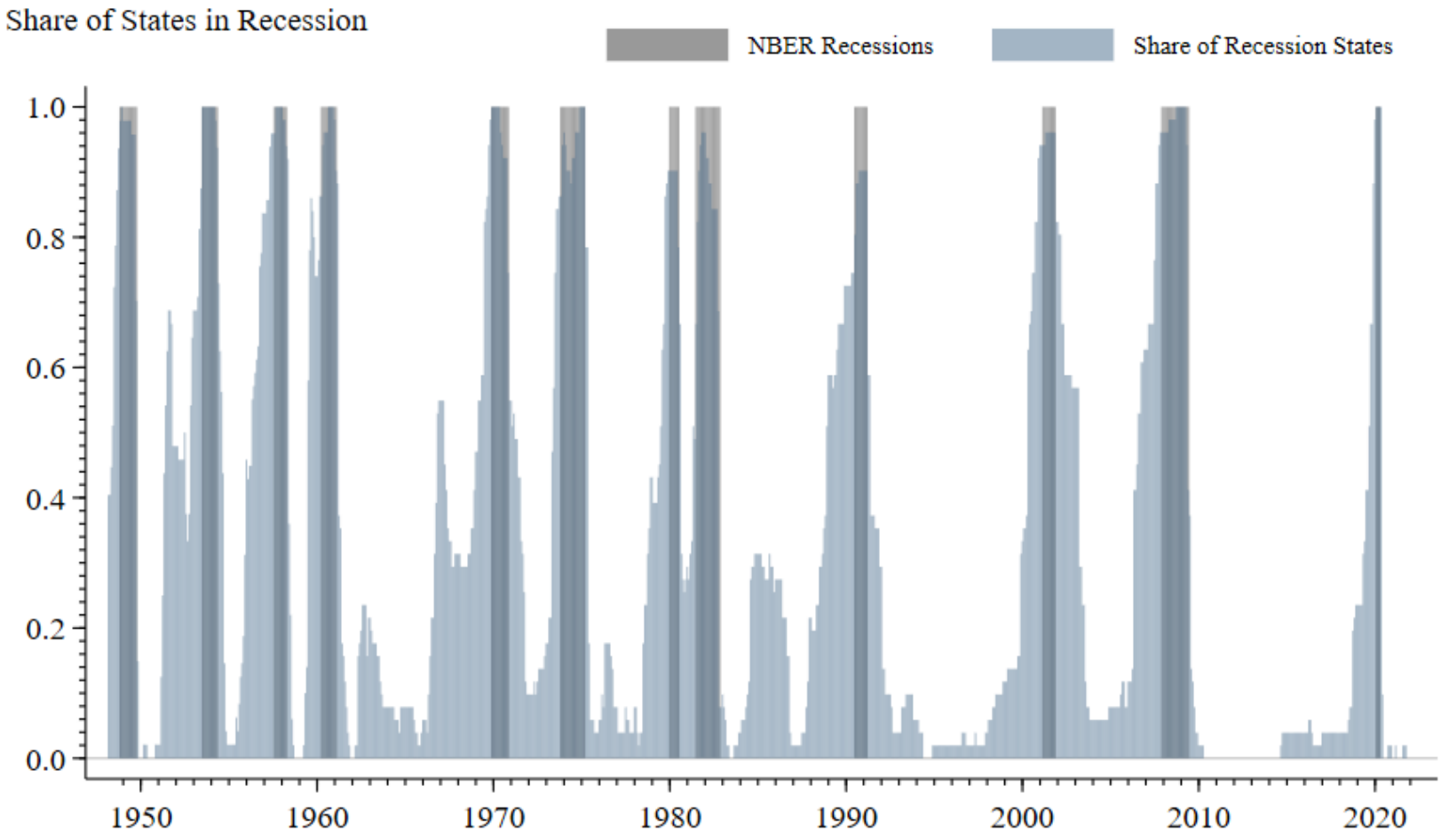
▸ Back

Recovery Pace: National Recoveries w/ CBUR Dates



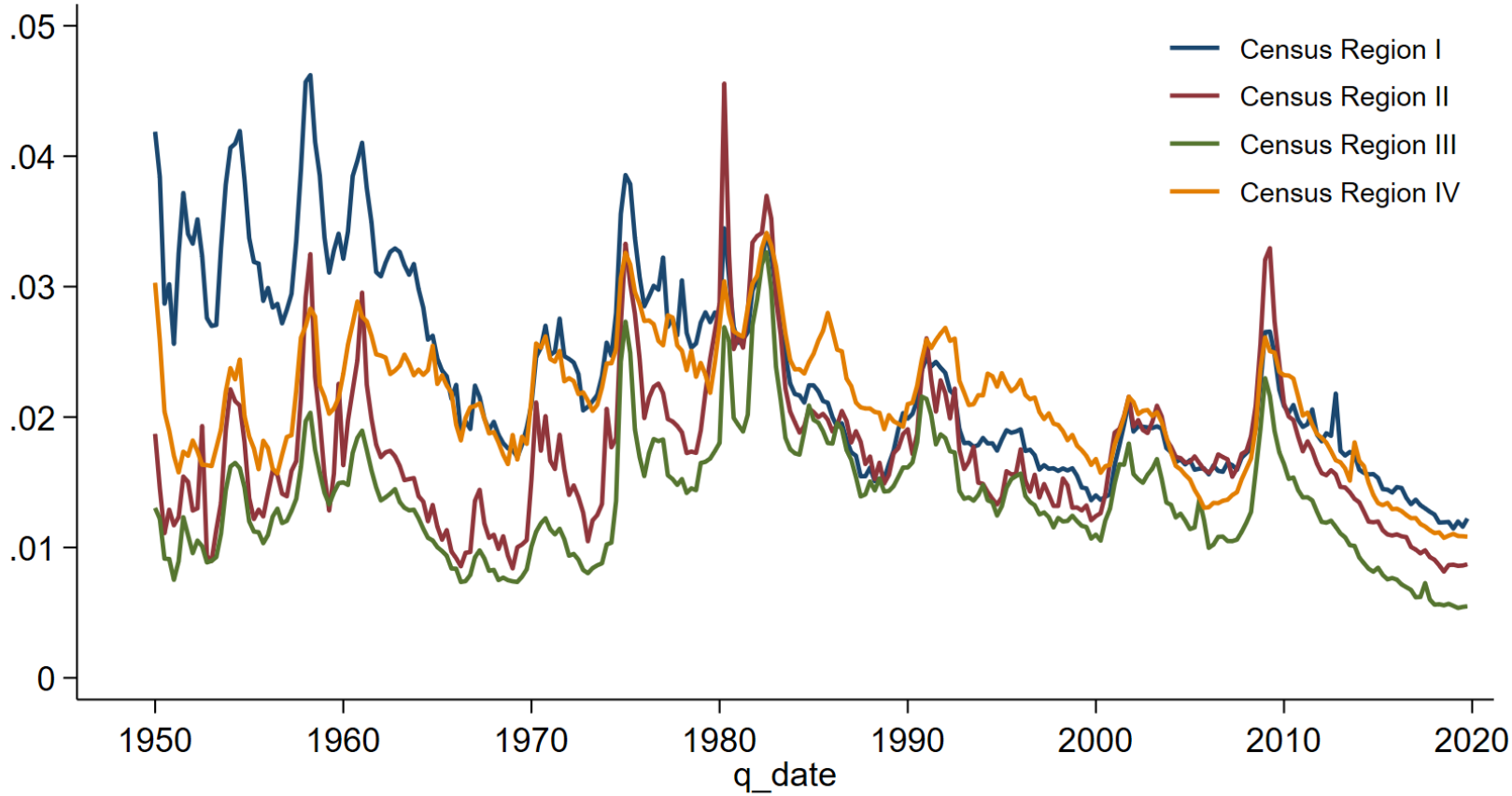
▸ Back

Recession Dating: State-level w/ CBUR Dates



▸ Back

Unemployment by Census Regions



Census Region I: CT, ME, MA, NH, RI, VT, NJ, NY, PA.
Census Region II IN, IL, MI, OH, WI, IA, KS, MN, MO, NE, ND, SD.
Census Region III: DE, DC, FL, GA, MD, NC, SC, VA, WV, AL, KY, MS, TN, AR, LA, OK, TX.
Census Region IV: AZ, CO, ID, NM, MT, UT, NV, WY, AK, CA, HI, OR, WA.