The Lost Generation? Labor Market Outcomes for Post Great Recession Entrants
Jesse Rothstein

Online Appendix
April 2021, revised December 2024

Note: This December 2024 revision of the appendix includes updated versions of appendix figures 2-7 that correct a minor data processing error. Changes are minor and generally invisible to the eye. A new Appendix C presents comparisons between the original figures distributed in the April 2021 version of this appendix and the corrected versions.

Appendix A. Data processing and definitions

The primary analysis concerns the cohort-age-year-state employment rate and mean log real wages for college graduates. Employment rates are estimated from the monthly Current Population Survey (CPS). I construct a synthetic panel from data from each month from 1979 through 2019, aggregated to the calendar year. Log wages are computed from the 1979-2017 Merged Outgoing Rotation Group (ORG) files from the monthly CPS, using an algorithm adapted from Center for Economic and Policy Research (2018; see also Schmitt [2003]). For those paid hourly, I use the higher of the reported hourly wage (without overtime, tips, or commissions) and the ratio of weekly earnings (which in principle include overtime, tips, and commissions) to usual weekly hours. For non-hourly workers, I use the latter. When necessary, usual hours are imputed using actual hours last week or the mean by gender and full time/part time status. Hourly wages are converted to real 2015 dollars using the CPI, set to missing if below $1 or above $200, and logged.

My sample for age-time-cohort decompositions consists of respondents born 1948-1997 (who turned 22 between 1970 and 2019) who are surveyed at ages 22-40 and who report being college graduates.

Individual respondents are assigned the age (and implied birth cohort) that they are on the survey date, so someone born on July 1, 1980 and surveyed in 2010 will be treated as age 29 in 2010 and in the 1981 birth cohort, if surveyed in the first half of the calendar year, and as 30 and in the 1980 cohort if surveyed in the second half. Before 1992, graduates are those with 16 or more years of completed schooling; afterward, it is those with Bachelors or graduate degrees.

In some specifications (e.g., in Figure 4), I identify the age, time, and auxiliary control coefficients only from pre-Recession cohorts. For these specifications, I estimate the decomposition using the subsample excluding post-1978 birth cohorts, who turned 22 after 2000. I use the resulting coefficients to form predictions for the subsequent cohorts if the relevant cohort effects were zero, then estimate these cohort effects as the average difference between observed outcomes and the predictions.

My unemployment measures are the state unemployment rates, averaged over the calendar years. When assigning the entry unemployment rate, I use the rate in the state of current residence in the year when the respondent was 22 years old. State unemployment rates are not available prior to 1976; for earlier entrants, I use the national unemployment rate.

All analyses are weighted by the sum of CPS sample weights in the cell.

Appendix B. Additional specifications

I present a range of additional specifications and models in the appendix. I discuss them in turn.

Educational attainment

I explore here changes across cohorts and within cohorts across ages in rates of degree attainment, with an eye toward understanding changes in my sample of young college graduates over time. Appendix Figure 1 shows the share of each cohort with some college or more (panel A) or with a bachelor's degree or more (panel B) by year and age, with dashed lines connecting successive observations on birth cohorts as they age.

Beginning with Panel A, we see that the share of each cohort that has some college or more has grown consistently across cohorts, save for a lull in the early 2000s and a return to rapid growth by around 2005. While there is growth across ages within cohorts, it is swamped by the growth across cohorts, such that at many times the share of 24-year-olds with some college exceeds that of 30-year-olds.

Patterns in degree attainment are more relevant for my analysis, which focuses on college graduates. Panel B of Appendix Figure 1 shows that the degree attainment rate for 22-year-olds is consistently less than half of what will be seen for the same cohort at age 30. For cohorts born in the 1960s, the great majority of degrees were earned by age 24. There was a substantial increase in post-24 degree attainment around the 1972 birth cohort (turning 24 in 1996; see Bound, Lovenheim, and Turner 2012), but still the great majority of degrees were awarded by 26. Since the mid-1990s, final attainment has grown substantially, with most of this growth in degrees awarded by age 24.[1] There is no indication either of a surge in BA attainment among Great Recession cohorts or of changes in the share of degrees earned at older ages.[2] Together, these facts make it plausible that age effects will absorb most of the bias in cohort mean outcomes averaged across various ages.

Main decompositions

Appendix Figures 2 and 3 show estimated age and time effects, respectively, from the main decomposition (2) for employment of college graduates. Age effects are normalized to zero at age 32, while year effects are normalized to zero in 2007. (As discussed in the main text, the cohort effects from this decomposition, shown as the solid line in Figure 3, are normalized to zero for both the 1984 and 2000 entering cohorts.)

[1] Post-24 degrees may be growing for the post-1988 birth cohorts (aged 30 in 2018 and later). It will be several years until we can fully measure this, and in the meantime it has little impact on my CPS samples.

[2] Panel A does show important increases in later-age college enrollment that are plausibly attributable to the Great Recession. This could create more serious sample selection biases in analyses of the population with any college.

Appendix Figure 4 reports estimates when I re-estimate my main age-time-cohort decompositions (shown here as the "base" specification) restricting the sample to those aged 22-27 and to entry cohorts between 1970 and 2014. This is a balanced cohort-age panel (at least at the end, though not at the beginning) rather than the balanced year-age panel used in the main analyses. Employment effects are largely unchanged from the main estimates, but wage trends are notably worse for post-Great-Recession entrants. Evidently, older cohorts saw poorer wage growth after the Great Recession than younger cohorts; when they are excluded, the post-2010 year effects rise and the corresponding cohort effects fall.

Appendix Figure 5 shows cohort effects on log wages from three different specifications. Cohort effects on employment from these same specifications are shown in Figure 4, and the specifications are described in the main text.

Appendix Figure 6 shows estimates of the medium-term scarring and excess sensitivity coefficients (θ_j and ϕ_j, respectively) from versions of my decompositions that are estimated on samples limited to pre-2000 entering cohorts. These are quite similar to the full sample estimates in Figure 6.

Sensitivity to mobility

In equation (3), I use the unemployment rate when the individual was 22, in the state of current residence, as a measure of labor market conditions when that individual entered the market. Interstate mobility between age 22 and the date of the CPS interview will make this a noisy measure of initial conditions. Further, any endogeneity of this mobility to local conditions could bias my results.

To address this, I use an instrumental variables strategy adapted from Schwandt and von Wachter (2019).[3] Construction of the instrument has three steps.

First, using pooled data from the 1980, 1990, and 2000 decennial census and from the 2001-2016 American Community Survey (one-year samples) public use microdata files, I estimate the number of college graduates born in each state *b* living in state *s* at age *a* ($22 \leq a \leq 40$), N_{bsa}. This is a long-run average, pooled across nearly forty years of data, so is not influenced by economic conditions relevant for any single cohort.

Second, I construct the average age-22 unemployment rate to which college graduates born in each state b and each entry cohort c would have been exposed. This average rate is then:

$$\widetilde{UR22}_{bc} = \frac{\sum_s UR_{sc} N_{bs,22}}{\sum_s N_{bs,22}}.$$

Third, for each state s, cohort c, and age a, I average the average birth-state exposure across all birth states, weighting by the share of age-a graduates in state s born in each state b:

[3] Schwandt and von Wachter (2019) use a double-weighting estimator to abstract from both endogenous mobility and endogenous attainment rates. I treat attainment rates as exogenous, using the selection correction discussed in the text to address changes in cohort composition with age.

$$\widetilde{UR22}_{sca} = \frac{\sum_b \widetilde{UR22}_{bc} N_{bsa}}{\sum_b N_{bsa}}.$$

Last, I interact this measure with age indicators $g_j(a)$ (j=1,…, 5) and use the results as instruments for the $D_a^j * UR_{sc}$ interactions in equation (3). Appendix Figure 7 shows the estimated medium-term scarring coefficients from the base OLS specification and using IV, first for equation (3) (in Panel A) and second for an augmented specification that also includes excess sensitivity controls (in Panel B). Coefficients are similar in IV as in OLS specifications.

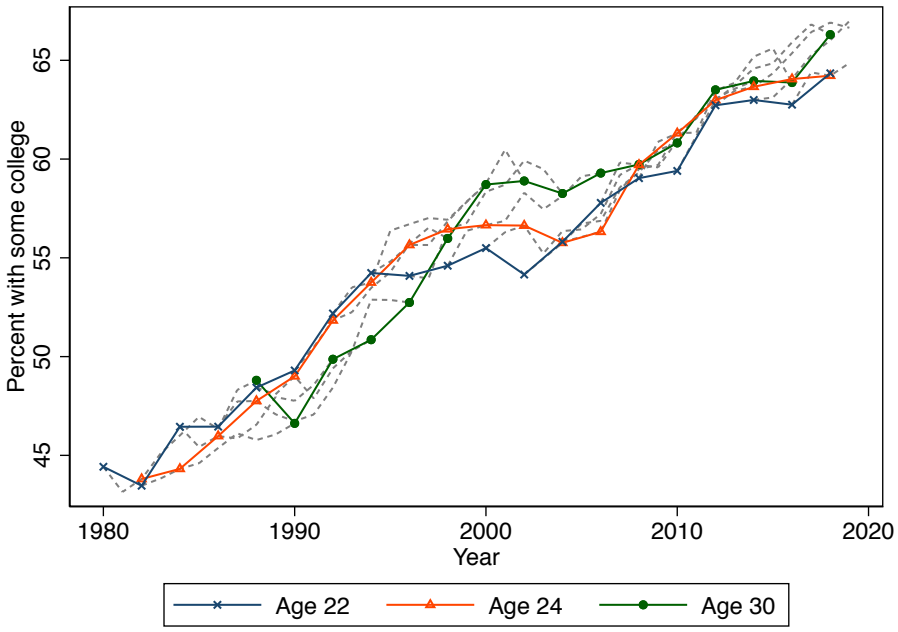
References

Center for Economic and Policy Research. 2018. CEPR ORG Version 2.3, Apr 10, 2018, http://ceprdata.org/wp-content/cps/programs/org/cepr_org_master.do.

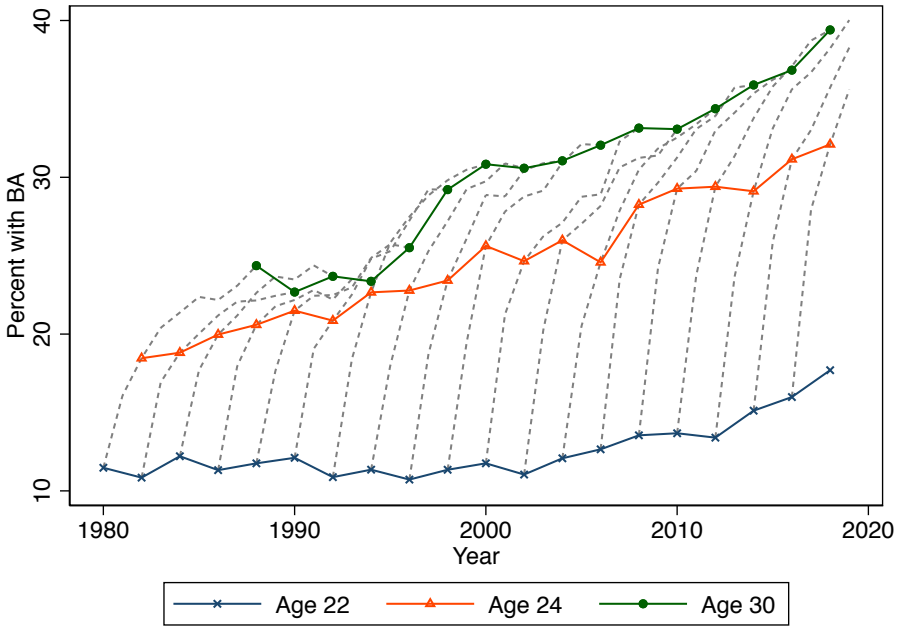
Schmitt, John (2003). "Creating a consistent hourly wage series from the Current Population Survey's Outgoing Rotation Group, 1979-2002." Center for Economic Policy Research technical paper, August.

Appendix Figure 1. Some college share and college graduate share by year and age

A. Some college share

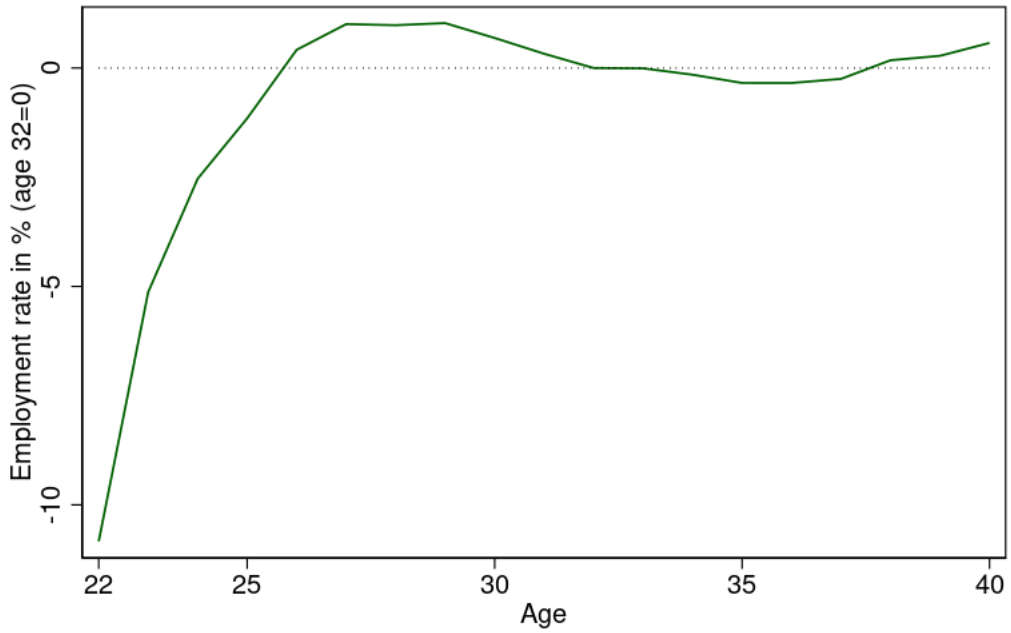


B. College graduate share



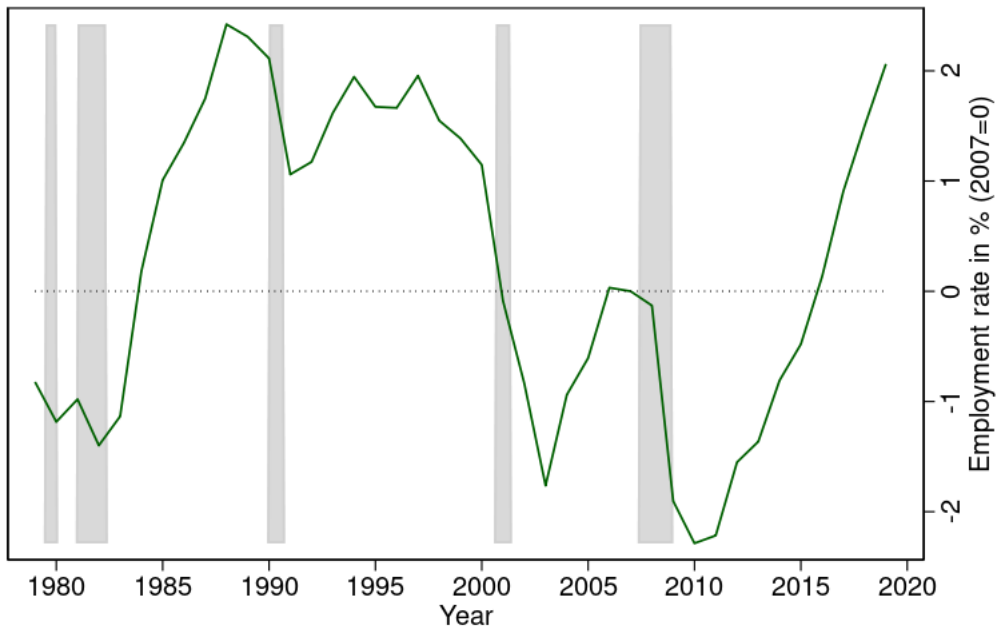
Notes: Each grey dashed line represents repeated cross-sectional averages for a single birth cohort, followed from 22 through 30 across successive CPS surveys. For example, the leftmost series includes those who were 22 in 1980, observed in the 1980-1988 CPS surveys. Only cohorts that are age 22 in even-numbered years are shown. Solid lines connect observations across cohorts at ages 22, 24, and 30. Crossings of these lines indicate that the cross-sectional ranking of these age groups by attainment is not stable over time.

Appendix Figure 2. Age effects from age-cohort and age-time-cohort decompositions of employment (*corrected, Dec. 2024)*



Notes: Figure shows estimated age effects γ_a on employment from equation (2), estimated on cohort-age-state-year cells constructed from CPS microdata samples of college graduates aged 22-40. The age-32 coefficient is normalized to zero; the model also normalizes the cohort effects to zero for both the 1984 and 2000 entering cohorts.

Appendix Figure 3. Time effects from age-time-cohort decomposition of employment (*corrected, Dec. 2024)*

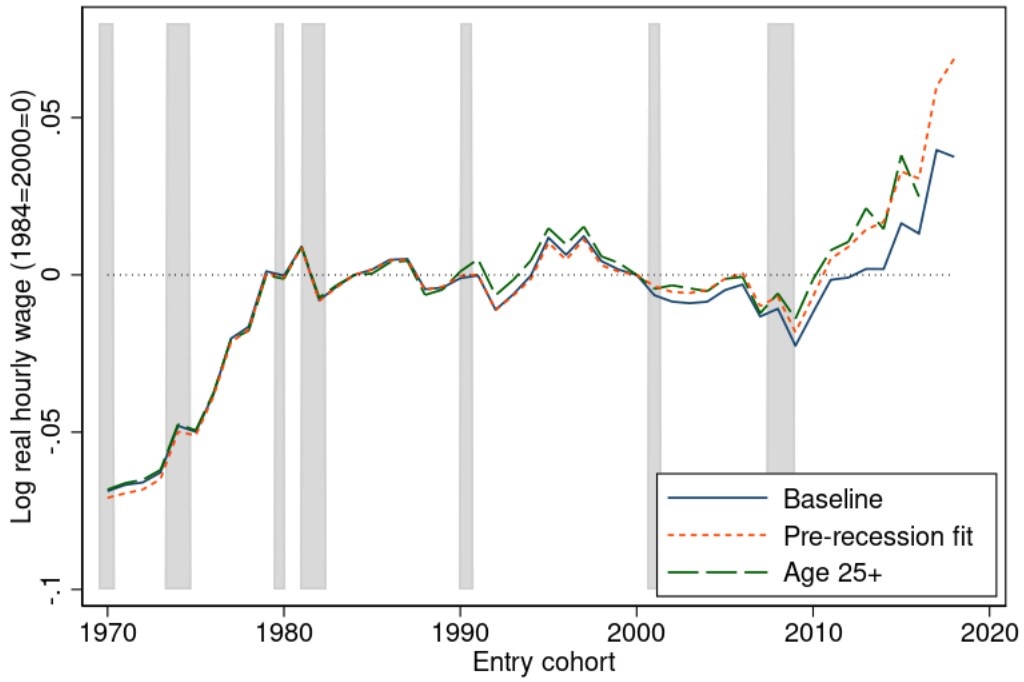


Notes: Figure shows estimated year effects β_t on employment from equation (2), estimated on cohort-age-state-year cells constructed from CPS microdata samples of college graduates aged 22-40. The 2007 coefficient is normalized to zero; the model also normalizes the cohort effects to zero for both the 1984 and 2000 entering cohorts. Recessions are indicated by shaded bars.

Appendix Figure 4. Cohort effects from balanced cohort-age specification (*corrected, Dec. 2024)*

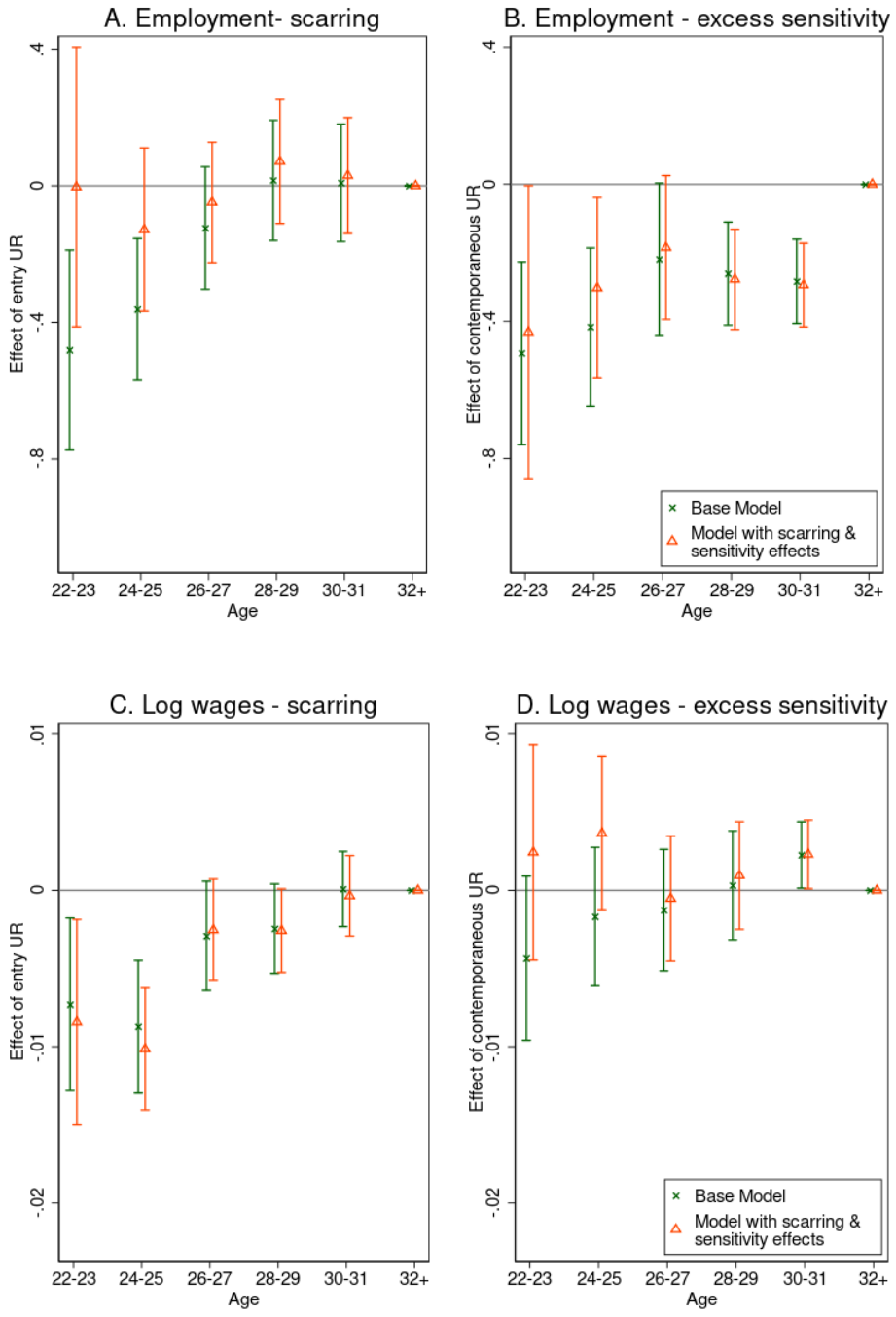
Notes: Base specifications repeat the age-time-cohort decompositions from Figures 3 and 5. "Balanced cohort-age" specifications repeat the same specifications, but limit the sample to ages 22-27 and the 1970-2014 entry cohorts. Estimated cohort effects for pre-1978 entrants from this specification are not shown.

Appendix Figure 5. Cohort effects on log wages of college graduates (*corrected, Dec. 2024)*



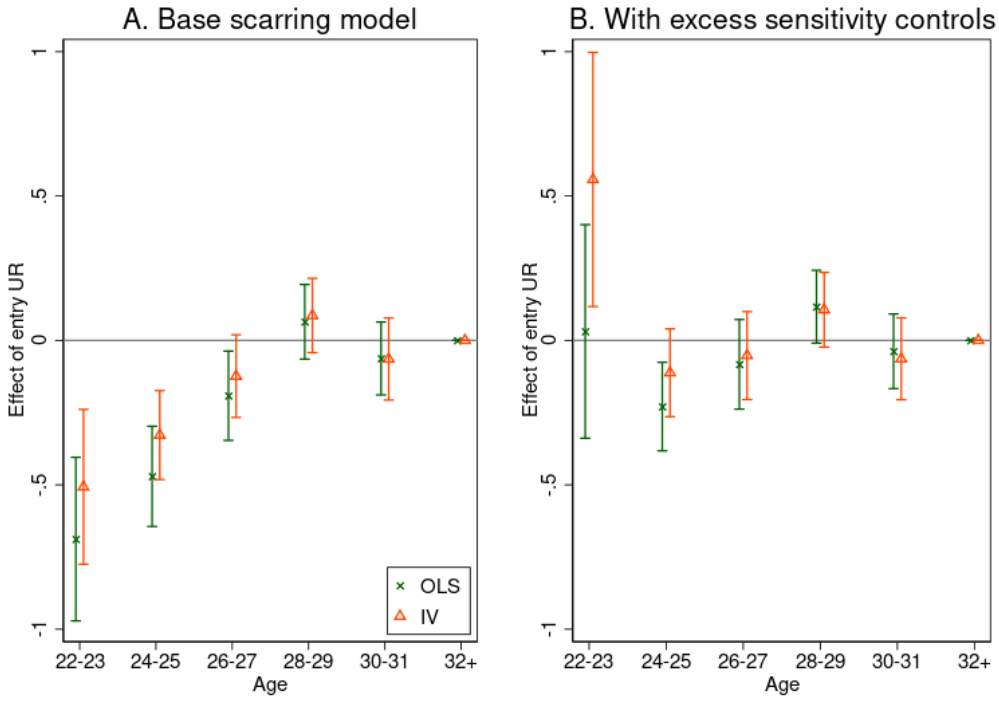
Notes: Figure shows cohort effects on log real hourly wages from three specifications, computed from CPS Outgoing Rotation Group samples of employed college graduates aged 22-40 and normalized to zero in 1984 and 2000. The "baseline" series repeats the "age-time-cohort decomposition" series from Figure 3. The "pre-recession fit" series estimates equation (2) using just cohorts that turned 22 in 2000 or earlier. Cohort effects for subsequent cohorts are estimated as the mean difference between observed log wages and predictions based on the estimated model. The "age 25+" model includes all cohorts (except the most recent entrants) in the sample, but removes observations when the cohort is less than 25 years old. The 2019 entrants are included in the estimation but their coefficients, which are very noisily estimated, are not plotted. Recessions are indicated by shaded bars.

Appendix Figure 6. Medium-term scarring and excess sensitivity estimates based on entrants in 2000 and earlier (*corrected, Dec. 2024)*



Notes: "Base model" derives from estimates of equations (3) (panels A and C) and (4) (panels B and D), and shows coefficients and confidence intervals, clustered at the state level, for ϕ_j and θ_j, respectively. Expanded model shows estimates from a model that includes both terms. Each is estimated on a subsample that excludes cohorts entering the labor market after 2000.

Appendix Figure 7. Instrumental variables estimates of medium-term scarring (*corrected, Dec. 2024)*



Notes: OLS estimates are those plotted in panel A of Figure 6. IV estimates are from a specification that instruments for $D_a^j * UR_{sc}$ with $D_a^j * \widetilde{UR_{sc,22}}$, as discussed in the text.

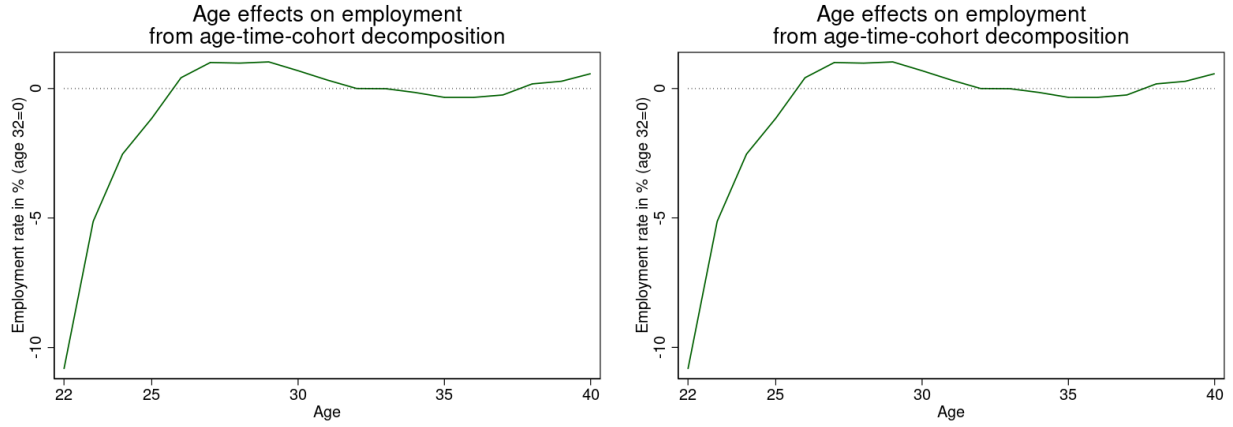
Appendix C. Erratum, December 2024

I recently discovered an error in the analysis files. Specifically, the analysis for the 2021 version of the paper used a faulty crosswalk between state Census codes and state FIPS codes. This crosswalk permuted the state codes in the West North Central division (Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota), leading Current Population Survey results for states in this region to be assigned to the wrong state (also in this division), and thus to the wrong unemployment rate series. States outside the West North Central division were unaffected.

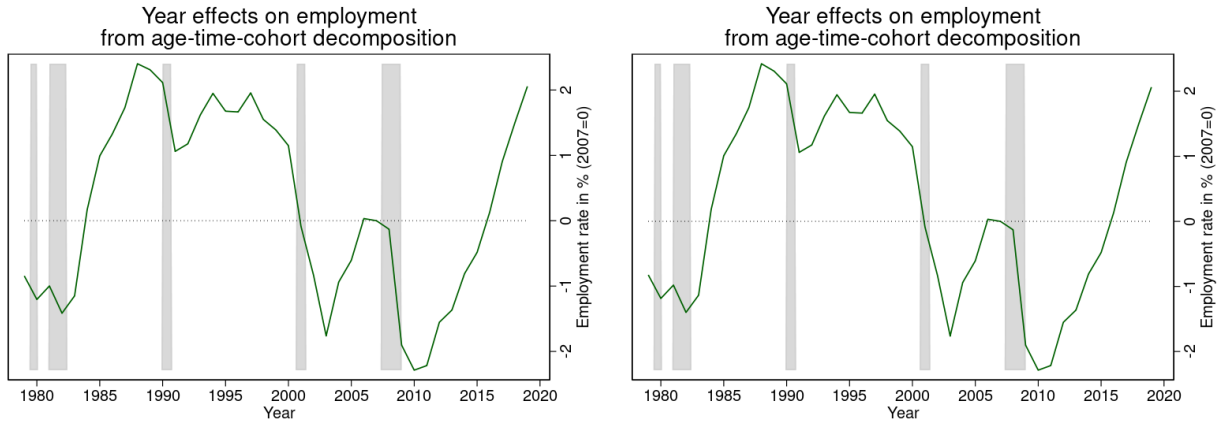
The figures above are based on corrected files that fix this mistake, using a corrected crosswalk. The correction has non-zero effects on Appendix Figures 2-7. Few are visible to the naked eye. In the interest of transparency, I present here side-by-side versions of the figures that enable readers to assess the changes. A separate erratum to be published on the JHR website provides a similar side-by-side comparison of the tables and figures in the paper itself. The replication archive will be updated as well to reflect the corrections.

Appendix Figure 2. Age Effects from Age-Cohort and Age-Time-Cohort Decompositions of Employment

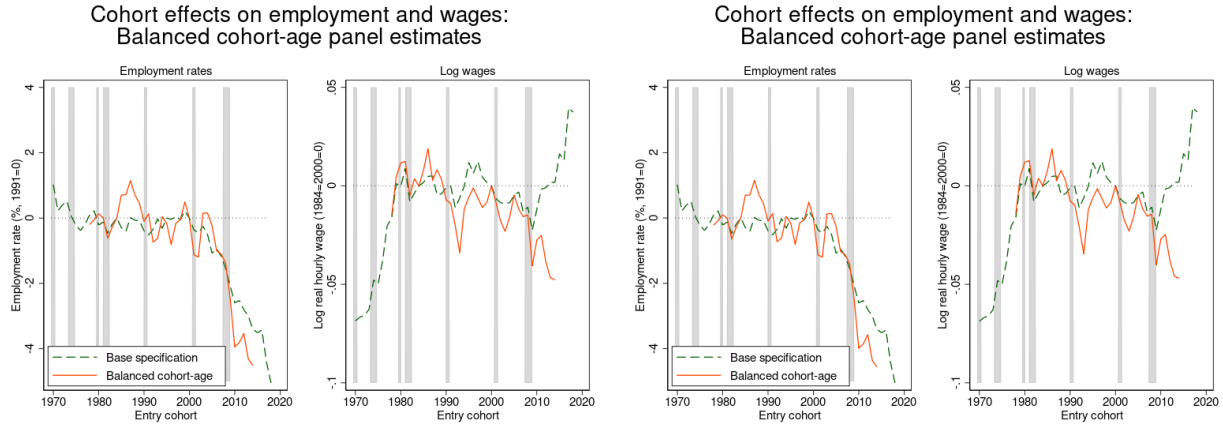
Left panel is published graph, reproduced from replication package; right panel is corrected graph



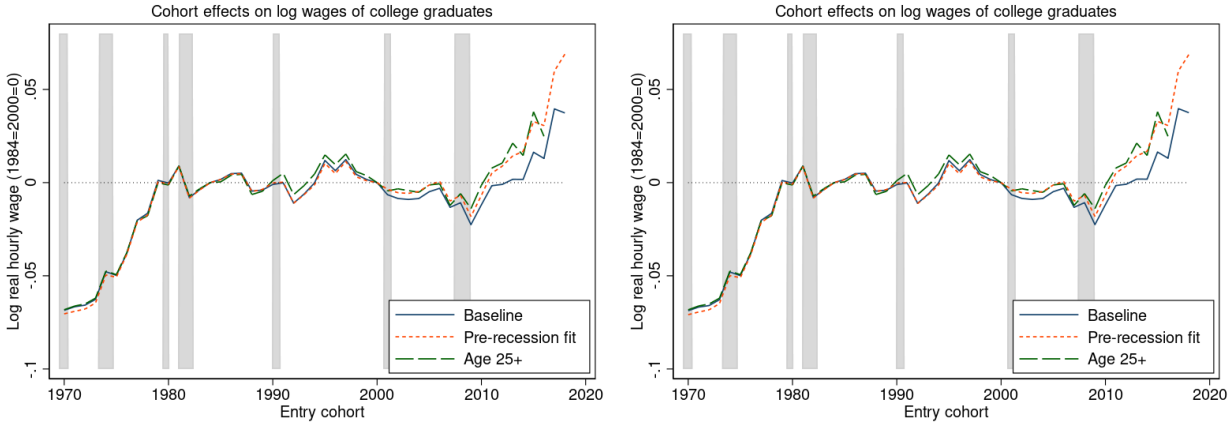
Appendix Figure 3. Time Effects from Age-Time-Cohort Decomposition of Employment
Left panel is published graph, reproduced from replication package; right panel is corrected graph



Appendix Figure 4. Cohort Effects from Balanced Cohort-Age Specification
Left panel is published graph, reproduced from replication package; right panel is corrected graph

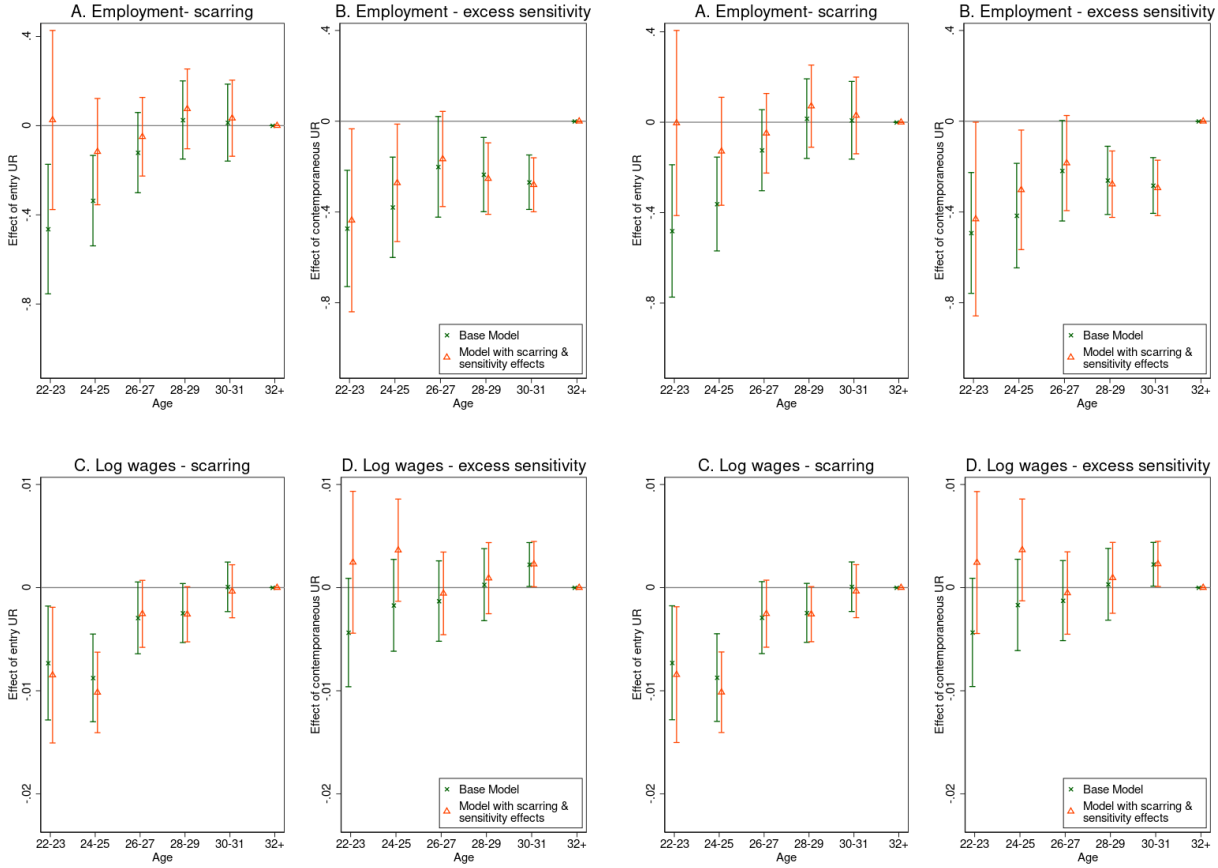


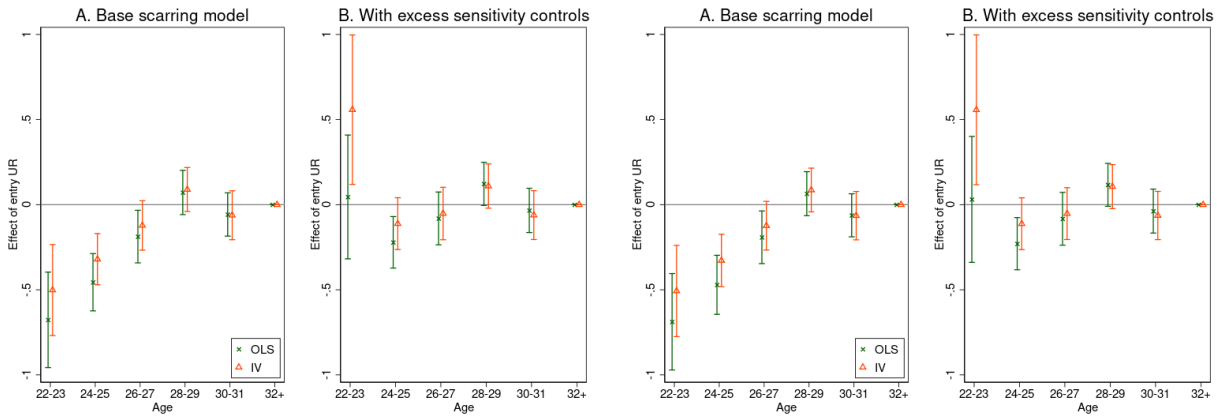
Appendix Figure 5. Cohort Effects on Log Wages of College Graduates
Left panel is published graph, reproduced from replication package; right panel is corrected graph



Appendix Figure 6. Medium-Term Scarring and Excess Sensitivity Estimates Based on Entrants in 2000 and Earlier

This Figure has four panels. The four panels on the left are from the published graph, reproduced from the replication package. The corrected panels are on the right.



Appendix Figure 7. Instrumental Variables Estimates of Medium-Term Scarring

Left two panels are published graph, reproduced from replication package; right panels are corrected graph