

Data Appendix for “Labor Market Polarization Over the Business Cycle”

Christopher L. Foote* Richard W. Ryan† Matthew Curtis‡

August 28, 2014

Abstract

This appendix describes the data, methods, and statistical programs used to construct the historical occupation-level employment and unemployment series in Foote and Ryan (2014). We also describe the occupational codes that are applied to 1976–2013 microdata from the Current Population Survey (CPS). Finally, we outline some standard statistical adjustments made to the occupation-level wage averages and worker flows generated from the CPS.

*Federal Reserve Bank of Boston. Email: Chris.Foote@bos.frb.org.

†University of Michigan. Email: RichRyan@umich.edu.

‡Federal Reserve Bank of Boston. Email: Matthew.Curtis@bos.frb.org.

The first three sections of this appendix discuss the construction of the historical employment and unemployment series in Foote and Ryan (2014). Section 4 describes some coding choices and statistical adjustments used for the CPS microdata. Because these topics are more standard they are covered in less detail. All Stata programs discussed below and their log files are available from the authors upon request.

1 Historical Sources for Occupational Employment and Unemployment

Table 1 in Foote and Ryan (2014), reproduced here as Table A1, shows how the historical occupational designations used by the Census Bureau and Bureau of Labor Statistics (BLS) can be grouped consistently into four theory-based classifications, following a suggestion by Jaimovich and Siu (2013). The four classifications are: nonroutine cognitive (high skill), routine cognitive (middle skill), routine manual (middle skill), and nonroutine manual (low skill). Table A2 lists the contemporary data sources used for various time periods, with the sources for employment levels shown in Panel A. From the late 1940s to mid-1959, occupational employment data were printed in the Census Bureau's *Monthly Report on the Labor Force* (Current Population Reports Series P-57). Before 1958, occupational employment is available only for the first month of each quarter (January, April, July, and October). As discussed below, we take this data structure into account when seasonally adjusting the final employment series. Beginning in January 1958, employment levels become available monthly and in July 1959, the employment data begin to appear in the BLS's *Employment and Earnings* publication. For months beginning in January 1983, Machine-readable data on occupation-level employment are available on the BLS website (www.bls.gov).¹

Panel B lists sources for the unemployment rates. Seasonally adjusted occupation-level data on the total number of employed and unemployed persons, the size of the labor force (the sum of employment and unemployment) and unemployment rates are available for 1958q1 through 1981q4 in BLS Bulletin 2096, a retrospective compilation of CPS data (U.S. Bureau of Labor Statistics 1982).² That publication is our source for 1958-1981 unemployment rates, which are available in seasonally adjusted form. For reasons discussed below, the 1957 unemployment rates come from the March 1967 *Employment and Earnings* issue from BLS. For unemployment rates from 1982q1 through 2013q4, we use unemployment rates generated from CPS microdata, as described below.

¹Though they are not used in this the paper, the BLS website also has monthly occupation-level data on unemployment rates from January 2000 on.

²Thanks are due to Ryan Michaels for pointing us to this resource, which also includes monthly data.

2 Constructing Historical Employment Levels

The following programs use the raw data listed in Panel A of Table A2 to construct consistent occupational employment series from 1947q3 to 2013q4.

2.1 The `occ_final.do` program

To construct the data on employment levels, we first aggregated the original source data into the four Jaimovich-Siu groups.³ This collection and aggregation is performed by the program `occ_final.do`. As Table A2 notes, the quarterly employment levels before 1958 are based on only one month per quarter: January, April, July, or October. The `occ_final.do` program also imputes data for one quarter that is not available in the source data: 1957q2. The quarterly values for this month are backed out by using annual average data for 1957, along with the three quarterly datapoints that are available for that year.⁴

2.2 The `js_quarterly.do` program

The dataset created by `occ_final.do` includes quarterly data for the four main occupational groups, but several adjustments are required before the dataset can be used in a cyclical study. The program `js_quarterly.do` performs these adjustments and is described below.

Imputing data for 1953q3. Like employment for 1957q2, occupational employment for 1953q3 does not appear in the relevant source data. Unlike the values for the later quarter, however, occupational employment for 1953q3 cannot be backed out from published annual averages for 1953, which were never published anywhere as far as we could determine. We therefore impute the 1953q3 data for each of the four employment series with a “trend-plus-residual” method. Specifically, the log of each series is regressed on a constant, a linear trend and dummies for the second, third, and fourth quarters. Estimation is performed on a sample that ends before 1971q1 due to a classification break in the quarter (described more fully below). Letting x_t^i denote the level of employment in quarter t for $i \in \{\text{nonroutine cognitive, routine cognitive, routine manual, nonroutine manual}\}$, the separate regressions for the four employment categories are specified as

$$\ln x_t^i = \beta_1 + \beta_2q_2 + \beta_3q_3 + \beta_4q_4 + \beta_5\text{trend} + \varepsilon_t^i. \quad (\text{A1})$$

Figure A1 displays the estimated residuals from this regression ($\hat{\varepsilon}_t^i$) in blue. The residuals for 1953q3, $\hat{\varepsilon}_{1953q3}^i$, are constructed by linearly interpolating the four residual series and are

³The original source data and the “first-line” programs that handle them are available upon request.

⁴The annual average data come from the May 1960 *Employment and Earnings Annual Average Supplement*.

depicted with red circles. Using these estimated residuals, the imputed employment levels for 1953q3 are

$$\hat{x}_{1953q3}^i = \exp \left(\ln \hat{x}_{1953q3}^i + \hat{\varepsilon}_{1953q3}^i \right).$$

The blue line in Figure A2 depicts employment in the four occupational groups from 1947q3 to 1970q4. The 1953q3 values, marked with red circles, are imputed using the procedure described above and are the only points in our dataset not directly generated from published data.

Excluding 14- and 15-year-olds from early data. The BLS began to exclude 14- and 15-year-olds from its published employment series beginning in 1967. Previously published data included those young workers, so we had to remove estimated employment of 14/15-year-olds from the early data to be consistent with post-1967 data. Estimates of early 14/15 employment are constructed by the program `agg1415.do` and read into `js_quarterly.do`. The basic strategy of `agg1415.do` is to first estimate a total employment-to-population (epop) ratio for 14/15-year-olds, and then separately estimate the share of 14/15 employment in each of the four occupational categories. Combining this information with the number of 14/15-year-olds in the population generates the estimates of 14/15 employment in each employment group.

Constructing the numerator of the 14/15 epop ratio requires an estimate of aggregate employment for the young workers. Fortunately, when BLS began to exclude 14/15 employment from its overall employment totals in 1967 it also began publish 14/15 employment separately.⁵ While the level of 14/15 *population* is not available from BLS, we can approximate 14/15 population with the currently published level of 16/17 population eight quarters ahead.⁶ The total epop ratio for 14/15-year-olds is modeled as:

$$\begin{aligned} & \log \left(\frac{\text{Employment}_{14-15}}{\text{Population}_{14-15}} \right)_t \\ &= \alpha + \underbrace{\beta_1 \log \left(\frac{\text{Employment}_{16-17}}{\text{Population}_{16-17}} \right)_t + \beta_2 \log (\text{unemployment rate})_t}_{\text{Business Cycle Effect}} \\ &+ \underbrace{\beta_3 \log \left(\frac{\text{Employment}_{16-17}}{\text{Population}_{16-17}} \right)_{t+8}}_{\text{Cohort Effect}} + \varepsilon_t. \end{aligned}$$

This regression model assumes that the current-quarter 14/15 epop ratio depends on the epop ratio for 16- and 17-year-olds in the current quarter, as well as the current unemployment rate and the 16/17 epop ratio eight quarters ahead. Including the current 16/17 epop ratio

⁵As noted below, BLS published 14/15 employment through 1982.

⁶We therefore ignore international migration, deaths, and other issues.

helps pick up contemporaneous business-cycle effects that are not captured by the aggregate unemployment rate. Including the 16/17 epop ratio from eight quarters in the future picks up cohort effects, because this ratio is generated by the same workers who are aged 14 and 15 in the current quarter. While 14/15 employment is available only from 1967 through 1982, the model can construct predicted values from 1948q1 onward because the right-hand-side variables go back that far. When we ran this model using OLS, the R^2 was high (.91) and all three of the main explanatory variables were significant at high levels (p-values of zero to three decimal points). Adding a linear trend raised the R^2 by less than 0.01 and caused both the trend and the unemployment rate to enter insignificantly (though they were jointly significant). We therefore used the model without the trend for parsimony.

The next step is to estimate the shares of 14/15 employment in each of the four occupational categories. Fortunately, the BLS disaggregated by occupation when it published 14/15 employment from 1967 through 1982. We therefore regress the level (not logs) of occupational shares in overall 14/15 employment on the aggregate unemployment rate and quarterly dummies. Because the aggregate unemployment rate extends back to 1948q1, we can construct estimates of 14/15 shares in each of the four occupational groups from 1948q1 through 1966q4.⁷ Finally, to extend the estimates back through 1947q3, we had to approximate the beginning of two series. First, the 16/17-year-old epop ratio is not available in 1947q3 and 1947q4. The year-ahead values are used in their place; for example, the value for 1947q3 is assumed to equal the value from 1948q3. Second, to get unemployment rates for 1947q3 and 1947q4, we used the 1948q1 value for both.

The effects of the 14/15 corrections appear in Figure A3. Not surprisingly, we find that most 14/15-year-olds worked in the low-skill, nonroutine manual group. Indeed, these young workers accounted for a nontrivial portion of employment in the low-skill category; some unreported calculations indicate that 14/15-year-olds accounted for between 3.6 and 6.2 percent of uncorrected low-skill employment before 1967.⁸ The figure also shows that the 14/15 correction generates a smoother series for low-skill employment over the 1967 break (denoted with a vertical line). The smoothness gives us some confidence that our correction is the right size.⁹

Spanning 1971q1 and 1983q1 changes in occupational classifications. As sug-

⁷In the share regressions, we also experiment by including linear trends. These trends made virtually no difference to the resulting estimates of 14–15 employment, so we omitted them from the model.

⁸Shares of 14/15-year-old employment among the other three occupational groups were 0.8 percent or smaller.

⁹A final note on the 14/15 correction: As noted above, BLS redefined the age range for published employment totals from 14+ to 16+ starting in 1967. Yet BLS reports for a given month typically list year-ago data defined in a comparable way, so we were able to use the BLS’s estimate of 16+ employment for 1966 that is available in a 1967 employment report. Thus, for statistical purposes our age-break occurs at the beginning of 1966, not 1967.

gested by Table A1, BLS occupational classifications changed little from the 1940s to the 1970s. Major changes did occur in 1971q1 and 1983q1, and both of these changes generated “seams” in our occupational series.¹⁰ To splice the data across these breaks, we model the four log employment levels with a VAR before each break, and then project fitted values across the breaks. For example, to splice the time series before 1971, a VAR process in log employment of the four occupational groups is estimated using data before 1971q1. Each of the four equations in the VAR includes four lags, quarterly dummies, and a linear trend.¹¹ Based on this estimated VAR(4) process, a forecast for each of the four occupational series is made for 1971q1. The difference between this forecast and the log level of published employment in 1971q1 is then subtracted from the log of published data from 1971q1 onwards. Because the model is specified in natural logs, this method effectively shifts the level of employment from 1971q1 onward by a constant percentage. The process is repeated to span the break in 1983q1, with 1971-break-corrected data prior to that quarter used to estimate a second VAR(4) with the same specification as before. We only need to perform two splices because BLS publishes consistent occupational data that begin in 1983. There have been a number of classification changes after 1983, most importantly in 2003 when the 2000 classification system is introduced. Yet BLS has essentially performed the required splice for us by combining employment levels across more finely disaggregated occupations in their published post-1983 data.

Figure A4 shows the results of our two break corrections. The most significant effect of the 1971 splice occurs for the low-skill nonroutine manual workers (Panel D), when the splice effectively shaves off a discrete increase in published employment for that year. The 1983 correction has more significant effects for all but the high-skill workers in Panel A.

The goal of the VAR adjustments is to move workers into consistent occupational categories, but nothing constrains the *sum* of employment in the four categories to equal total nonagricultural employment before and after each splice. It is gratifying to note that this condition comes close to being met anyway. The solid lines in Figure A5 depict total nonagricultural employment reported by BLS during the postwar era. The solid green line depicts employment in nonagricultural *industries* from 1948q1 to 1982q4. Starting in 1983q1, BLS makes available employment for nonagricultural *occupations*, which is shown by the solid red line. The two dashed lines in Figure A5 present sums of the four occupation-specific employment levels from 1947q3 through 2013q4. The dashed blue line shows the sum before

¹⁰The breaks actually occur in the first months of 1971 and 1983, with the introduction of the 1970 and 1980 occupation codes, respectively. But we use quarterly data for all statistical models, so our breaks occur in the first quarter of each of those years. Also, when introducing the 1970 codes in 1971, BLS noted that the switch to the 1970 codes was the most comprehensive change in occupational coding since 1940 (Bregger 1971). This fact helps explain why no apparent seams exist before 1971, and it provides some justification for our decision not to perform any splices before then.

¹¹The VAR is also run after the 14/15 correction described above.

any VAR adjustments are made, while the brown line shows the VAR-adjusted sum. The similarity of the two dashed lines after the first splice in 1971q1 indicates that the splices do not change the implied total of nonagricultural employment very much. And the close correspondence of the dashed lines with the solid lines suggest that this implied total is close to the BLS's current estimate of nonagricultural employment throughout the postwar era.

Seasonal adjustment. To seasonally adjust the four occupational employment series, we use something similar to a ratio–moving-average method. The main difference is that we allow the seasonal cycle to change at two exogenous dates. We first detrend the log levels of the four employment series with Hodrick–Prescott (HP) filters.¹² We then regress the resulting log deviations on three sets of quarterly dummies. The first set is not interacted with any other variable; the second set is interacted with an indicator set equal to one before 1958q1; and the third set is interacted with an indicator that equals one starting in 1983q1. The seasonal cycle is thereby allowed to change in 1958 and 1983. The first change is required because, as noted in Table A2, occupational employment is available only for the first month in each calendar quarter before 1958. The 1983 change is suggested by the apparent change in seasonality in that year for some of the four employment categories, most notably for low-skill, nonroutine manual employment (discussed in more detail below).

The residuals from the quarterly-dummy regressions constitute seasonally adjusted log deviations of employment from HP trends. Merging these residuals back to the original HP trends results in seasonally adjusted levels of employment. Figures A6 through A9 show final versions of the occupational employment levels in each of the four occupational categories. The solid black line in each figure depicts the data after the 1953q3 imputation and the VAR splices, but before any seasonal adjustment.¹³ The solid blue lines depict seasonally adjusted data using an HP filter with a smoothing parameter of $\lambda = 1,600$. The dashed red line seasonally adjusts with an HP filter with $\lambda = 100,000$, the baseline choice in Foote and Ryan (2014).

All four figures indicate that the choice of smoothing parameter makes little difference, with the slight exception of the seasonally cyclical routine manual workers early in the sample period (Figure A8). Additionally, Figure A9 confirms that allowing the seasonal cycle to change in 1983 is a particularly good idea for the low-skill, nonroutine manual category, as employment in this series becomes more seasonal in 1983. One potential explanation for the increase in low-skill seasonality is that the 1983 classification change moved some seasonal workers who used to be classified as routine manual into the low-skill category.¹⁴

¹²Our baseline specification uses a smoothing parameter λ equal to 100,000. We discuss robustness to different values of λ below.

¹³Summing these data generates the brown dashed line in Figure A5.

¹⁴Figure A4 showed that the 1983 VAR adjustment raises employment in the routine manual group while it reduces employment in the low-skill nonroutine manual group. This type of splice would be required to

For another check on the seasonal adjustment procedure, we can use some seasonally adjusted data available in BLS Bulletin 2096, which is a retrospective data collection published by BLS in 1982 (U.S. Bureau of Labor Statistics 1982).¹⁵ Figure A10 compares our seasonally adjusted employment data (using an HP trend with $\lambda = 100,000$) with data from 1958-1981 that appears in Bulletin 2096.¹⁶ There is a close correspondence between the two sources, with the lone exception consisting of a jump in published BLS employment for low-skill workers in 1971q1. Our first VAR adjustment smoothed out this seam, induced by the 1971q1 classification change, but it remained in all the official BLS documents we found until the introduction of the 1980 classifications in 1983, at which point BLS ceased publication of pre-1983 occupational employment levels.

3 Constructing Historical Unemployment Rates

In addition to occupation-level employment levels, BLS Bulletin 2096 also includes unemployment levels and unemployment rates over the same period (1958–1981). The unemployment rates we use are seasonally adjusted by BLS and collected by the program `alldata.do`, which also reads in seasonally adjusted unemployment rates from various issues of *Employment and Earnings*. The latter rates are available because the BLS allows its seasonal adjustment factors to change over time, so that seasonally adjusted data from recent time periods can change as time elapses. To keep published data up-to-date, the BLS typically published revised seasonally adjusted data for various series in an *Employment and Earnings* issue early in each calendar year. These rate updates are not preferred to the entire 1958–1981 series that is available on a consistent basis in Bulletin 2096. However, the seasonal update in the March 1967 issue of *Employment and Earnings* includes some unemployment rates that are not available in Bulletin 2096, specifically, the unemployment rates from 1957. We therefore used unemployment rates from the March 1967 *E&E* for unemployment rates in 1957 and those from Bulletin 2096 for the years 1958-1981.¹⁷ These early rates appear as the solid

“undo” a 1983 classification change that moved routine manual workers into the low-skill group. An obvious question concerns the *business-cycle* seasonality of any routine manual workers that BLS moved into the low-skill group. If employment for these workers were highly sensitive to the business cycle, then the cyclicity of low-skill employment would rise mechanically with the addition of these workers. However, the bottom row of Figure 6 in Foote and Ryan (2014), shows that the cyclical properties of the two groups does not change much after 1983: employment of routine-manual workers remains cyclically sensitive while employment of low-skill workers remain acyclical.

¹⁵We also use Bulletin 2096 to construct unemployment rates, as described below.

¹⁶Note Figure A10 depicts an apples-to-apples comparison, because the BLS retrospective was published after 1967. Thus the relevant data in Bulletin 2096 excludes 14- and 15-year-olds, as does our seasonally adjusted data.

¹⁷The BLS probably published unemployment rates from 1958 on in BLS Bulletin 2096 to preserve consistency with its publication of monthly occupational employment levels, which begin in that year. Recall that the employment data are available only for one month in each quarter before 1958. We are unaware of any historical BLS publication with occupational unemployment data before 1957.

blue lines in Figure A11, where the legend explicitly references the relevant sources of data.

Figure A11 also depicts unemployment rates we constructed from CPS microdata. To generate these rates, we first added the variable OCC1950 to the CPS microdata. OCC1950 was designed by Matthew Sobek to be a consistent categorization of occupations in Census IPUMS microdata.¹⁸ As its name suggests, OCC1950 is based on the 1950 system of occupational classifications, so once it is added we can generate occupational unemployment rates that are consistent with the earlier rates in Bulletin 2096. The program that merges the earlier unemployment rates with those implied by OCC1950 and CPS microdata is `ur.do`.

Figure A11 shows that published rates and the microdata-generated rates are close to each another during the 1976–1981 period of overlap, though it is clear that the early rates are seasonally adjusted while those constructed from microdata are not. The `ur.do` program exploits this discrepancy to construct seasonal adjustment factors for the later microdata that are consistent with the seasonal adjustment inherent in the published 1976–81 data. The program regresses the log of the published unemployment rates on the log of the rates generated by the microdata in the overlap period. Quarterly dummies are also included in this regression, generating a set of additive seasonal adjustment factors. These factors are used to seasonally adjust the later unemployment rates based on OCC1950 and CPS microdata. The resulting time series of unemployment rates are depicted in the top panel of Figure 5 in Foote and Ryan (2014).

4 Occupational Coding and Adjustments in CPS Microdata

4.1 Occupational classifications in Autor (2010) and Acemoglu and Autor (2011)

Sections 3, 4 and 5 of Foote and Ryan (2014) exploit CPS microdata by disaggregating middle-skill data by industry and by examining the labor-market transitions and participation decisions of individuals or demographic groups. The occupational designations in these sections is somewhat different than those for the historical analysis, in that they adopt the occupational grouping in both Autor (2010) and Acemoglu and Autor (2011). This classification system is based on 10 consistent occupations specifically designed to be applied to CPS microdata. These 10 occupations are based in turn on the 1990 classification system as redefined by David Dorn.¹⁹ The 10 occupations are grouped into the three skill classes as follows:

- **High skill:** Managers; Professionals; Technicians.

¹⁸For a description of OCC1950, as well as an alternative variable based on the 1990 occupational classification system, see Meyer and Osborne (2005). For a description of the IPUMS data, see Ruggles et al. (2010).

¹⁹The 1990 codes with which Dorn worked are similar to the 1980 codes, which are listed in Table A1 and Table 1 of Foote and Ryan (2014).

- **Middle skill:** Sales; Office and Administration; Production, Craft, and Repair; Operators, Fabricators, and Laborers.
- **Low skill:** Protective Services; Food Preparation; Building and Grounds Cleaning; Personal Care and Personal Services.

This three main groups above can be easily expanded to approximate the four Jaimovich-Siu groups by splitting the middle-skill group into a routine cognitive group, consisting of sales and office/administration, and a routine manual group, consisting of precision/craft/repair and operators/fabricators/laborers.

Figure A12 compares the employment levels implied by this system to the employment levels based on published sources developed above. As expected, the cyclical properties of occupation-level employment from the two sources are similar, though employment levels built up from the microdata experience seams when classification systems change. As indicated by the graphs of various unemployment and transition rates in Foote and Ryan (2014) and this appendix, however, these seams do not cause much trouble when calculating *rates* as opposed to *levels* in CPS microdata. A classification change that causes an employment-level increase of five percent could make a big difference to the measured cyclical properties of employment levels for a particular group if left uncorrected. But this addition would probably have a negligible effect on unemployment or transition rates, particularly if the workers added to the group behaved similarly to the workers who had already been included.

4.2 Time-aggregation adjustments for EU flows

In the modern search-and-matching literature, Shimer (2005) was among the first to point out a time-aggregation problem when measuring worker flows. The CPS is a point-in-time survey, so it measures a respondent's labor-market state at a specific time in a given month. If the respondent loses his job after one survey date but then finds a new one before the next month's survey date, his flow through the unemployment pool will be missed by the point-in-time CPS. A similar situation arises when a person reports unemployment in one month but then finds and subsequently loses a job before the next month's survey. This person's flow through employment will also be missed by the CPS. In both cases, the rate at which workers move between employment and unemployment impacts the measurement of the flow rate in the opposite direction.

The current state-of-the-art correction for this problem is described in Shimer (2012) and Elsby, Hobijn, and Şahin (2013). This correction uses an eigenvalue transformation to convert a 3×3 matrix of measured CPS transition rates among employment, unemployment, and nonparticipation into a matrix of continuous flow rates. Given these flow rates, expected transition rates over a single month can be calculated as if the time-aggregation problem

does not exist. We performed this transformation on our flow rates, though our inability to ascertain the skill class for nonparticipating workers meant that our transition matrix did not include flows through nonparticipation and was thus 2×2 rather than 3×3 . In practice, we found that the eigenvalue method generated corrected transition rates that were nearly identical to those generated by assuming that employed workers who lose their jobs have, on average, half a month to find a new job before they are surveyed again in the CPS. Likewise, workers who find jobs have, on average, half a month to separate again before they the next month’s survey. This assumption results in a particularly simple time-aggregation correction, because an employed worker who separates will tend to find another job before the next survey at a probability equal to one-half of that month’s job-finding rate. Similarly, the time-aggregation correction for unemployed workers finding jobs will involve one-half of the overall job-separation rate for that month.²⁰

4.3 Demographic adjustments for worker flows and hourly wages

Adjustments for EU and UE transitions. Two graphs in Foote and Ryan (2014) include adjustments intended to hold constant the demographic characteristics of groups of workers over time. The first demographic adjustment applies to the transition rates between employment to unemployment (EU and UE flows). The raw transition data for these flows come from matching workers between months in the CPS.²¹ A demographic correction is then applied to the EU flows (“job-separation rates”), and the corrected flows for middle-skill workers are depicted in Figure 10 of Foote and Ryan (2014). A demographic correction is also applied to UE flows that generate the economy’s overall “job-finding rate,” and important component of the unemployment-transition model presented in the paper’s Section 4.

The demographic adjustment for the EU and UE flows is based on a series of yearly logit regressions and is performed by the program `compadj5_find_sep.do`. This program models the probability of an EU or UE flow as dependent on a vector of demographic variables, which include a cubic in age and dummies for white, female, married, and educational attainment. The four possible educational categories are less than high school, completed high school, completed some college, or earned at least a college degree. Finally, quarterly dummies are also included.

The yearly regressions use matched data from January through December of each year.²² To be included in a yearly samples for the EU flows, the CPS respondent must be employed

²⁰The half-month assumption and its implicit correction are similar to some earlier work in Shimer (2005).

²¹Madrian and Lefgren (2000) describes the matching algorithm. Because the CPS is a monthly dataset, transition rates between labor market states are figured on a month-to-month basis. As is standard in the literature, we construct quarterly averages of these monthly rates for our analysis.

²²Occasionally, workers cannot be matched between months. This problem occurs in Dec. 1975/Jan. 1976, Dec. 1977/Jan. 1978, Jun. 1985/Jul. 1985, Sep. 1985/Oct. 1985, Dec. 1993/Jan. 1994, and May 1995/Jun. 1995 to Aug. 1995/Sep. 1995. See Shimer (2012, footnote 18) for further details.

in the first month of the match; whether the person is unemployed in the second month of the match (1 or 0) is the dependent variable.²³ Five separate samples are constructed based on occupation and industry in the first month of the match. These five samples use the high-skill, low-skill and middle-manufacturing, middle-construction and middle-other disaggregation described in Foote and Ryan (2014).

Separate logit regressions are estimated for these five groups, and the coefficients from the separate regressions are used to predict the probability a hypothetical person in a group transitions from employment to unemployment. This hypothetical person is constructed using demographic averages over the entire sample, 1976–2013. Specifically, the hypothetical person is based on the average participant in the labor force. The average age is computed over the entire sample. For the other regression variables, which are binary, the proportion of the entire sample is used. The hypothetical average person is constructed using final weights provided with CPS microdata. The quarterly dummies included in the regressions provide a demographically adjusted quarterly series, although this particular seasonal adjustment allows the seasonal cycle to vary across years and worker classifications.

The demographic adjustment for the UE transition rates uses the same procedure, though to be included in the samples a person must be unemployed in the first month of the match, and whether the person is employed in the second month of the match is the binary dependent variable. Again, separate logit regressions are run for each of the five industry-skill groups.²⁴

Adjustment for wage levels. A separate demographic adjustment is applied to the wage data in the second panel of Figure 14 in Foote and Ryan (2014). This adjustment is based on a demographic correction used by Haefke, Sonntag, and van Rens (2013) and proceeds as follows. Hourly wages in the CPS are typically defined as usual weekly earnings divided by usually weekly hours. Before this ratio is taken, we trim the highest and lowest 0.5 percentile of usual hours worked. The hourly wage variable is constructed by dividing the weekly earnings variable by the trimmed value of hours worked. Then the top and bottom 0.3 percentile of hourly wages are trimmed.

Wages are then compositionally adjusted in a two-step procedure. In the first step, individual log wages are regressed on the following observable characteristics: years of education, a quartic in experience, gender, race, and marital status.²⁵ The goal is to obtain parameters that reflect the direct effects of demographic characteristics, which are assumed to be the same regardless of the skill class of the worker. We can write the first-step regression as

$$\ln w_{ijt} = x'_{ij}\beta_t + \epsilon_{ijt},$$

²³Thus a zero in the second month corresponds to either employment or nonparticipation.

²⁴Both the EU and UE flows are demographically adjusted and then adjusted for time aggregation.

²⁵Experience is defined as age minus years of education minus six.

where w_{it} is the log hourly wage for worker i in skill group j at time t , x_i is a vector of worker-level demographic characteristics, β_t is a time-specific vector of coefficients, and ϵ_{it} is a residual. Note that the demographic influences are *not* assumed to be specific to skill groups, as the β s are indexed by t , not jt .

In the second step, we use the estimated β s and the within-group means of the demographic variables to construct skill-specific average wages that are not influenced by changes in the demographic composition of the skill group over time. Let w_{jt} denote the average *unadjusted* hourly wage for skill group j at time t , and $\bar{\beta}$ denote the average of the coefficient vector over all time periods. The log of the demographically adjusted wage is calculated as:

$$\ln \widehat{w}_{jt} = \ln w_{jt} - (x_{jt} - \bar{x}_j)' \bar{\beta},$$

where x_{jt} is the average demographic characteristics for skill group j at time t and \bar{x}_j is the average of these characteristics for group j across all time periods. The calculation constructs a demographically adjusted wage for each quarter and skill group, $\ln \widehat{w}_{jt}$, by subtracting off the time-varying demographic effects for each group. These effects are modeled as deviations from skill-group means, so the calculation assures an equality between the mean of the adjusted average wage $\ln \widehat{w}_{jt}$ and that of the unadjusted average wage $\ln w_{jt}$ over the sample period.

To see how this correction works intuitively, consider a group for which the average wage is rising over time primarily because an increasing percentage of workers are made up of college graduates. The effect of college attainment on earnings is estimated in the first-stage regression using individuals from all skill groups and reflected in the estimated β s. The rising share of college-educated workers in the given skill group will show up as rising values of the relevant element of the demographic vectors for that skill group, x_{jt} . Because the demographic effects are purged from the adjusted wage \widehat{w}_{jt} , changes in the demographically adjusted wage will stem only from developments orthogonal to changes in educational attainment and other demographic trends.

References

- Acemoglu, Daron, and David Autor. 2011. “Skills, Tasks and Technologies: Implications for Employment and Earnings.” In *Handbook of Labor Economics*, eds. Orley Ashenfelter and David E. Card, vol. 4. Elsevier.
- Autor, David H. 2010. “The Polarization of Job Opportunities in the U.S. Labor Market: Implications for Employment and Earnings.” A paper jointly released by The Center for American Progress and The Hamilton Project. Available at <http://economics.mit.edu/files/5554>.
- Bregger, John E. 1971. “Revisions in Occupational Classifications for 1971.” *Employment and Earnings* 17(8): 5–8. February.
- Elsby, Michael W.L., Bart Hobijn, and Ayşegül Şahin. 2013. “On the Importance of the Participation Margin for Labor Market Fluctuations.” Federal Reserve Bank of San Francisco Working Paper No. 2013-05. Available at <http://www.frbsf.org/publications/economics/papers/2013/wp2013-05.pdf>.
- Foote, Christopher L., and Richard W. Ryan. 2014. “Labor Market Polarization over the Business Cycle.” Forthcoming in *2014 NBER Macroeconomics Annual*.
- Haefke, Christian, Marcus Sonntag, and Thijs van Rens. 2013. “Wage Rigidity and Job Creation.” *Journal of Monetary Economics* 60(8): 887– 899.
- Jaimovich, Nir, and Henry E. Siu. 2013. “The Trend is the Cycle: Job Polarization and Jobless Recoveries.” Duke University Working Paper. Available at <https://sites.google.com/site/nirjaimovich/papers/JPJR.pdf>.
- Madrian, Brigitte C., and Lars John Lefgren. 2000. “An Approach to Longitudinally Matching Current Population Survey (CPS) Respondents.” *Journal of Economic and Social Measurement* 26: 31–62.
- Meyer, Peter B., and Anastasiya M. Osborne. 2005. “Proposed Category System for 1960–2000 Census Occupations.” U.S. Bureau of Labor Statistics, Office of Productivity and Technology. Working Paper no. 383. Available at <http://www.bls.gov/ore/abstract/ec/ec050090.htm>.
- Ruggles, Steven, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek. 2010. *Integrated Public Use Microdata Series (IPUMS): Version 5.0*, [Machine-readable database]. Minneapolis: University of Minnesota.

Shimer, Robert. 2005. “The Cyclical Behavior of Equilibrium Unemployment and Vacancies.” *American Economic Review* 95(1): 25–49.

Shimer, Robert. 2012. “Reassessing the Ins and Outs of Unemployment.” *Review of Economic Dynamics* 15(2): 127–148.

U.S. Bureau of Labor Statistics. 1982. Bulletin 2096. *Labor Force Statistics Derived from the Current Population Survey: A Databook*. September.

Theory-Based Occupational Group	Selected Census-Based Major Occupation Groups Used in Current Population Survey			
	1940 Groups	1970 Groups	1980 Groups	2010 Groups
Nonroutine Cognitive (High Skill)	Professional & Semi-Professional Workers Proprietors, Managers & Officials, excluding Farm	Professional, Technical & Kindred Workers Managers & Administrators, excluding Farm	Professional Specialty Technicians & Related Support Executive, Administrative & Managerial	Professional & Related Occupations Management, Business & Financial Operations Occupations
Routine Cognitive (Middle Skill)	Clerical & Kindred Workers Salesmen and Saleswomen	Clerical & Kindred Workers Sales Workers	Administrative Support, including Clerical Sales Occupations	Office & Administrative Support Occupations Sales & Related Occupations
Routine Manual (Middle Skill)	Craftsmen, Foremen & Kindred Workers Operatives & Kindred Workers Laborers, excluding Farm & Mine	Craftsmen & Kindred Workers Operatives, excluding Transport Transportation Equipment Operatives Laborers, excluding Farm	Precision Production, Craft & Repair Operators, Fabricators & Laborers	Production Occupations Transportation & Material Moving Occupations Installation, Maintenance & Repair Occupations Construction & Extraction Occupations
Nonroutine Manual (Low Skill)	Domestic Service Workers Service Workers, excluding Domestic	Service Workers	Service Occupations	Service Occupations

Table A1. CONSISTENT OCCUPATIONAL GROUPS IN THE CURRENT POPULATION SURVEY. Note: The mapping between the Census's major occupational groups and the four theory-based occupational groups is from Jaimovich and Siu (2013).

Panel A: Employment Levels			
Period	Source	Periodicity	Seasonally Adjusted?
1947q3–1957q4	Census Monthly Reports on the Labor Force (Current Population Reports, Series P-57)	Quarterly (Jan, Apr, Jul, Oct only)	No
1958m1–1959m6	Census Monthly Reports on the Labor Force (Current Population Reports, Series P-57)	Monthly	No
1959m7–1982m12	BLS Employment and Earnings	Monthly	No
1983m1–2013m12	Currently reported occupational-level data from BLS FTP site	Monthly	No
Panel B: Unemployment Rates			
Period	Source	Periodicity	Seasonally Adjusted?
1957m1–1957m12	March 1967 BLS Employment and Earnings	Monthly	Yes
1958q1–1981q4	BLS Bulletin 2096 (1982), Vol. 2	Quarterly	Yes
1976q1–2014q4	Implied quarterly unemployment rates from CPS microdata, coded to 1950 occupation codes	Quarterly	No

Table A2. ORIGINAL DATA SOURCES FOR EMPLOYMENT AND UNEMPLOYMENT DATA.

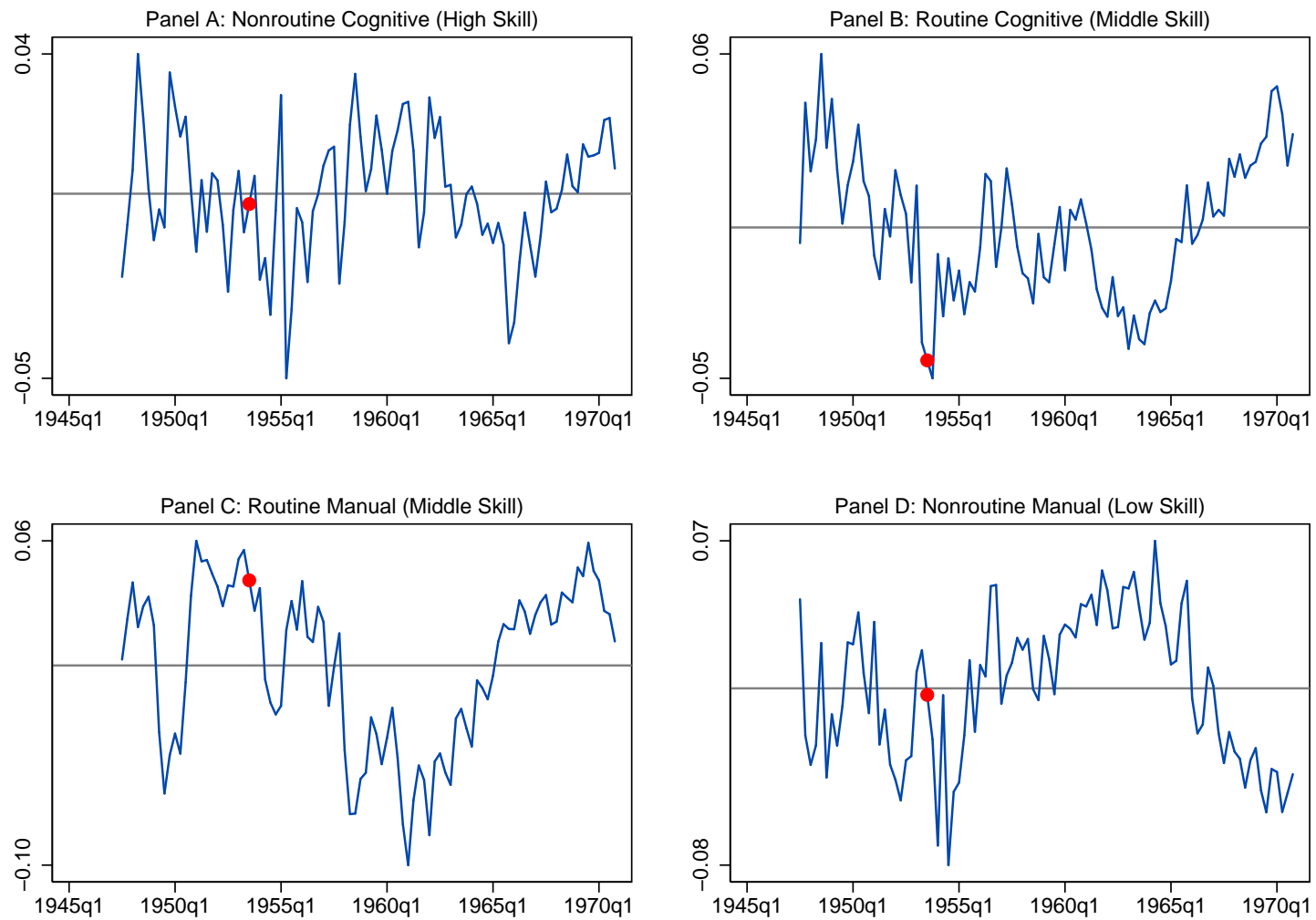


Figure A1. ESTIMATED RESIDUALS FROM REGRESSIONS OF EMPLOYMENT LEVELS USED FOR DATA IMPUTATION. Note: Regressions are specified as in Equation A1. The red circles depict linearly interpolated data for 1953q3.

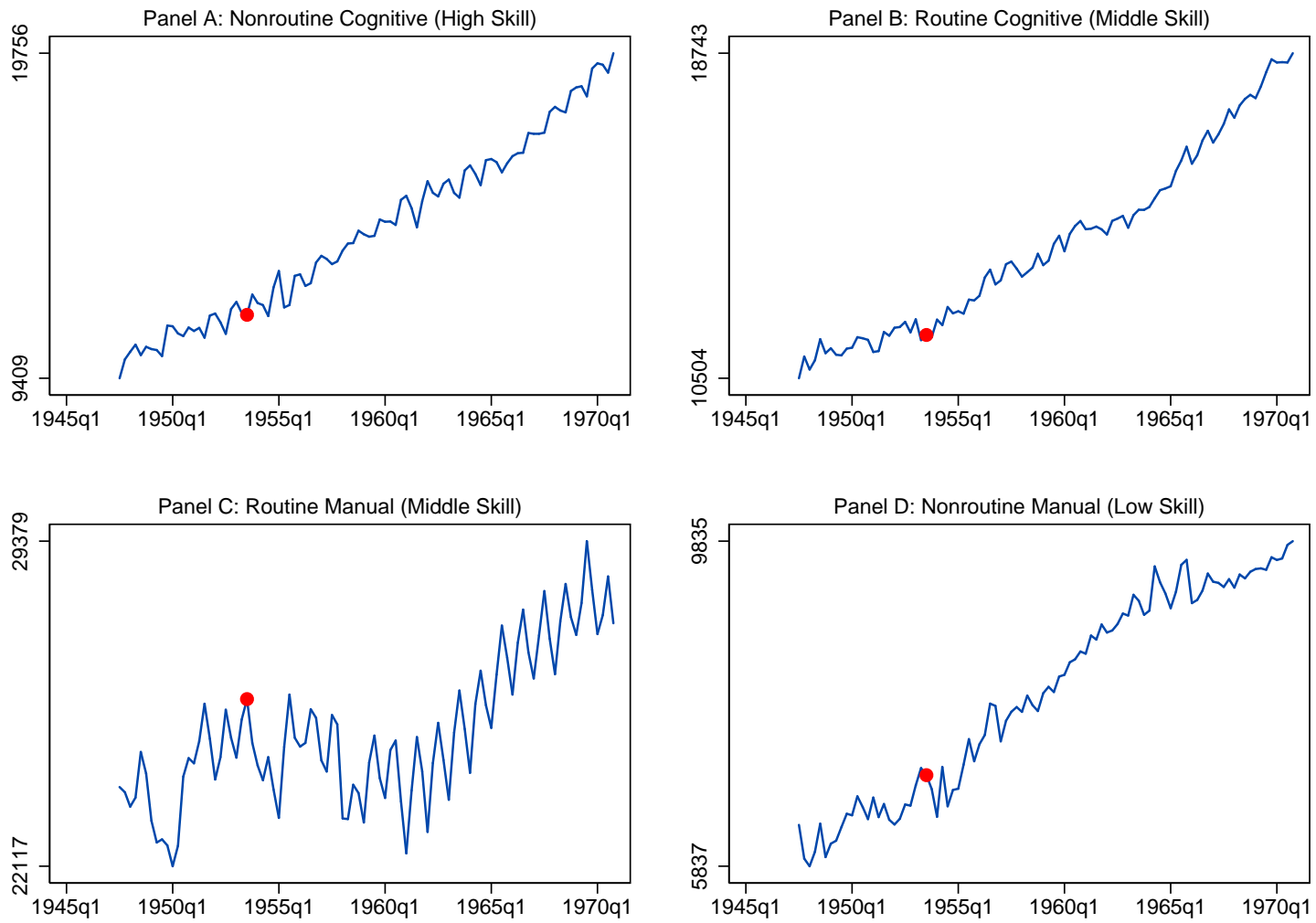


Figure A2. IMPUTED EMPLOYMENT LEVELS FOR 1953Q3. Note: Imputed data are depicted by red circles.

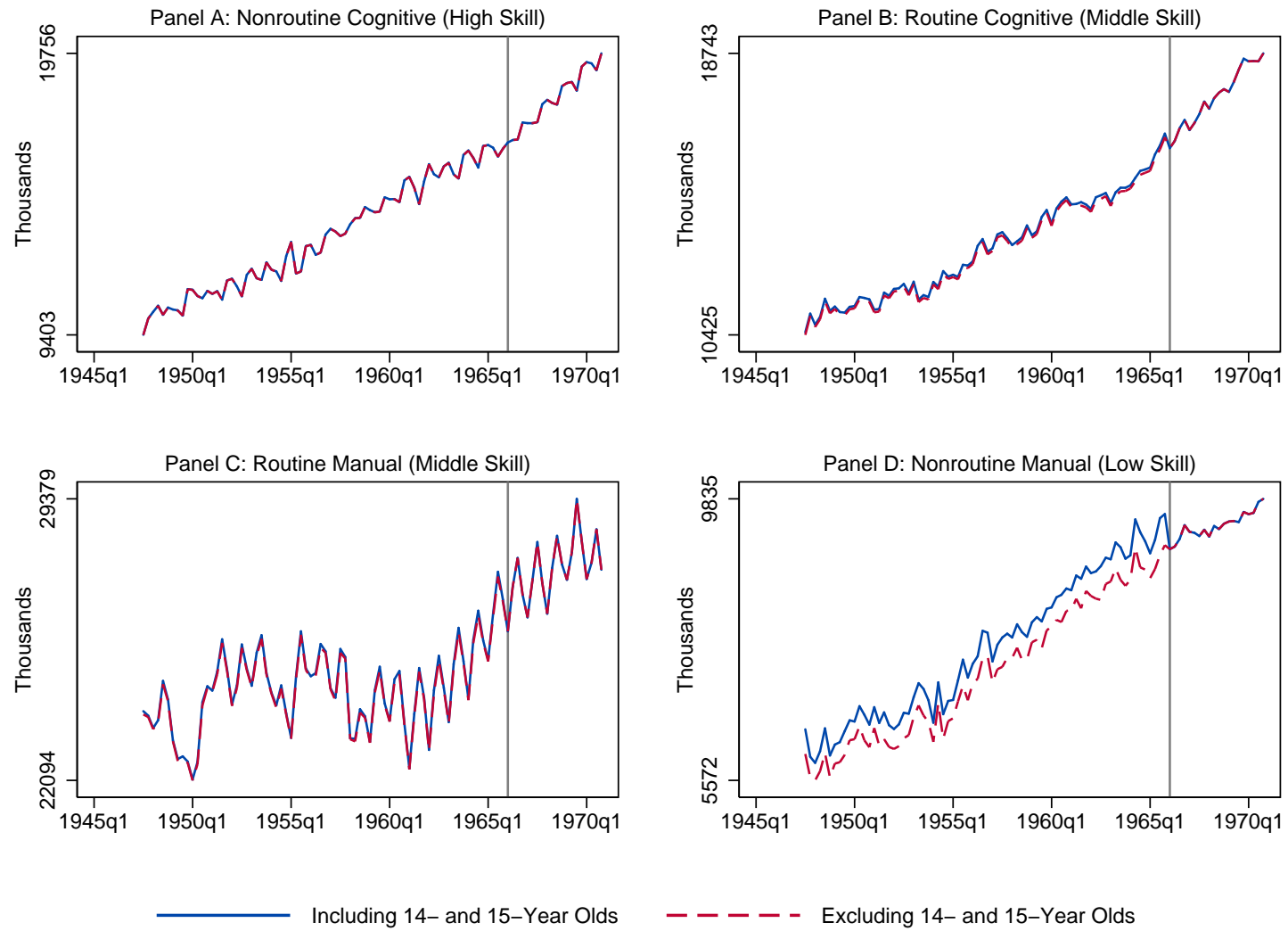


Figure A3. EFFECT OF REMOVING ESTIMATED EMPLOYMENT OF 14- AND 15-YEAR-OLDS FROM EARLY EMPLOYMENT DATA. Note: Vertical line denotes 1967q1, when BLS removed 14- and 15-year-olds from headline employment series.

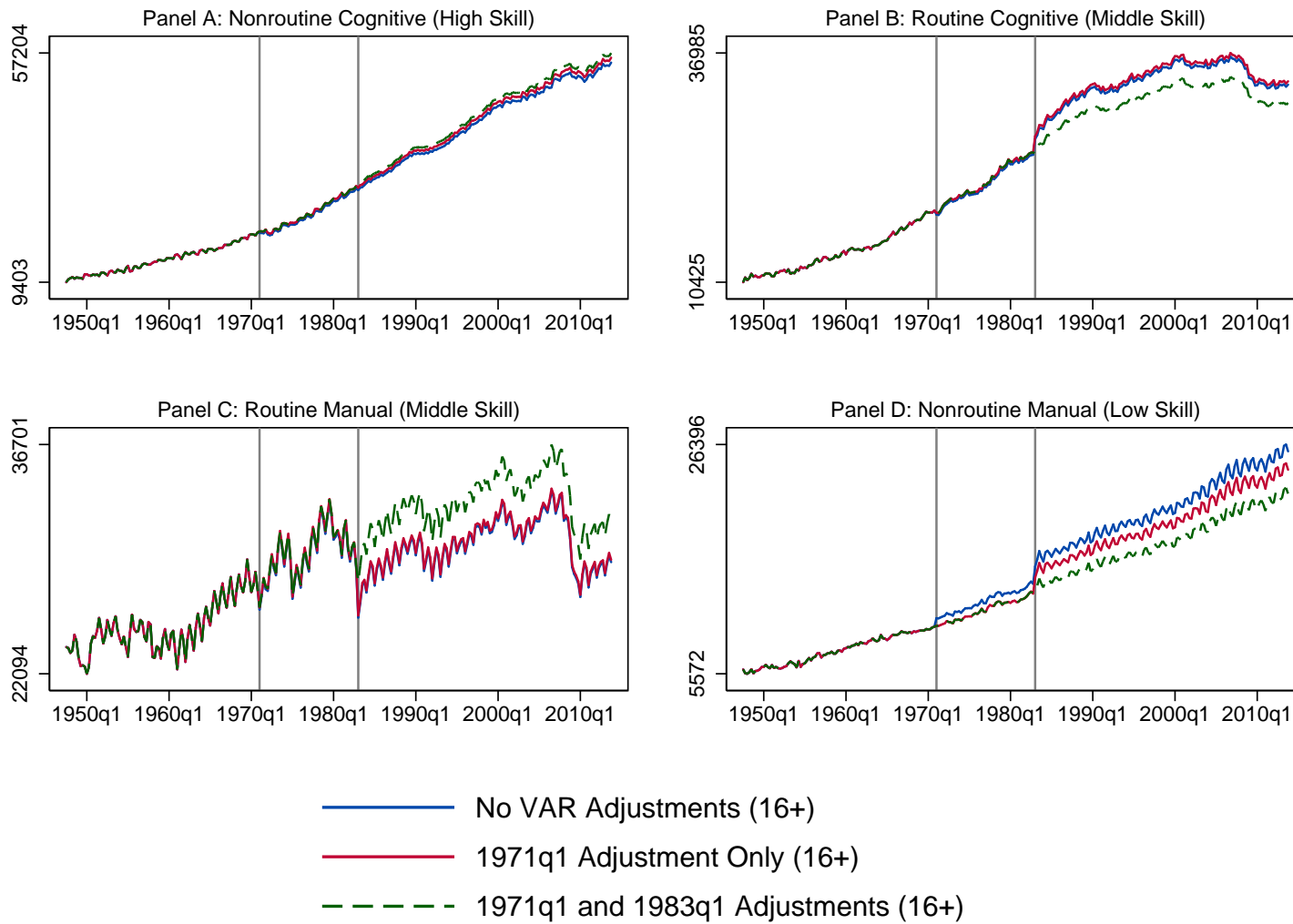


Figure A4. EFFECT OF ADJUSTMENTS FOR TWO OCCUPATIONAL CLASSIFICATION CHANGES. Note: Vertical lines denote 1971q1 and 1983q1, when the 1970 and 1980 occupational classification codes were introduced.

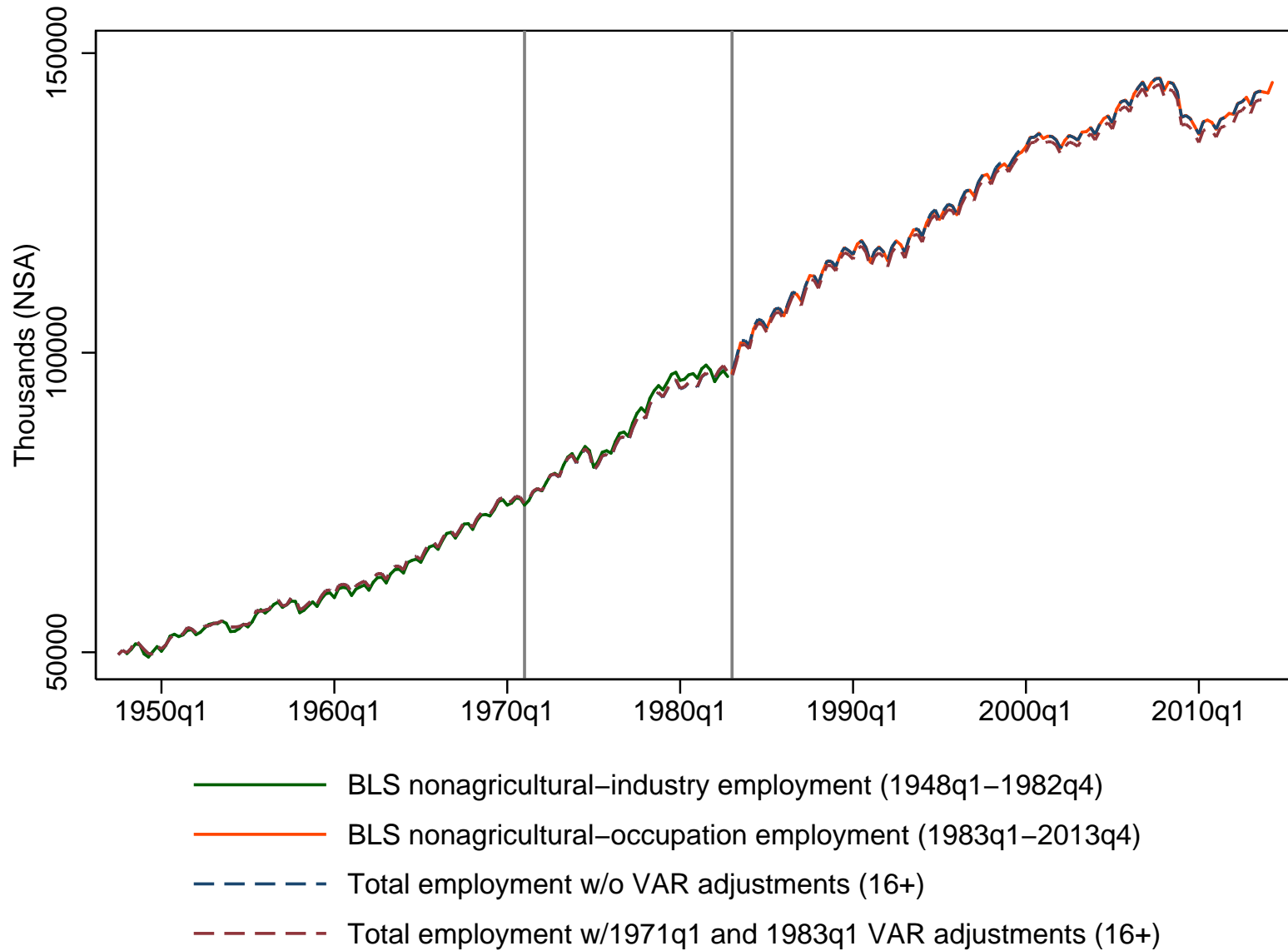


Figure A5. COMPARING SUM OF ADJUSTED OCCUPATIONAL EMPLOYMENT SERIES (DASHED LINES) WITH CURRENT BLS ESTIMATES OF TOTAL NONAGRICULTURAL EMPLOYMENT (SOLID LINES).

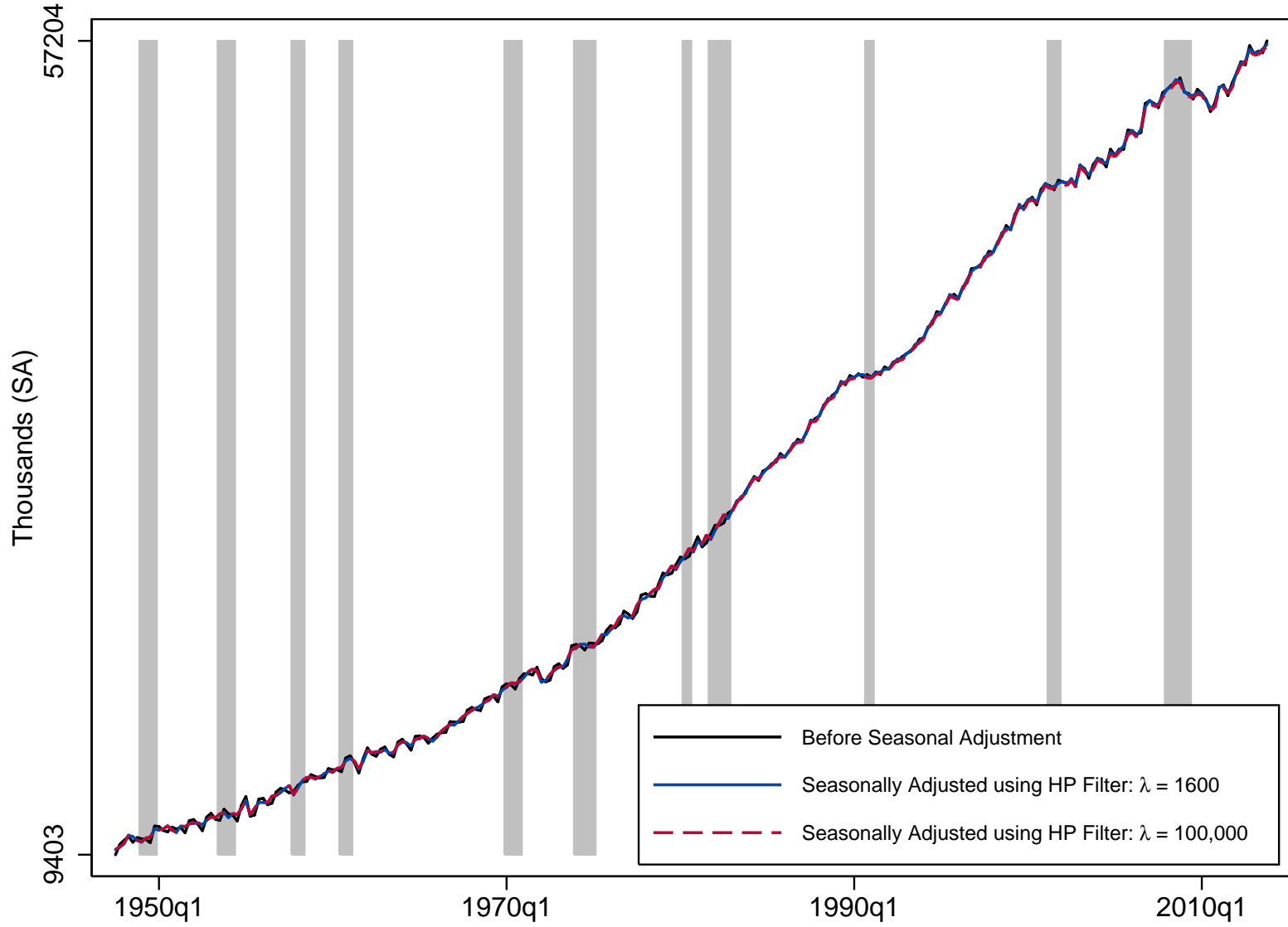


Figure A6. EMPLOYMENT OF NONROUTINE COGNITIVE WORKERS (HIGH SKILL).

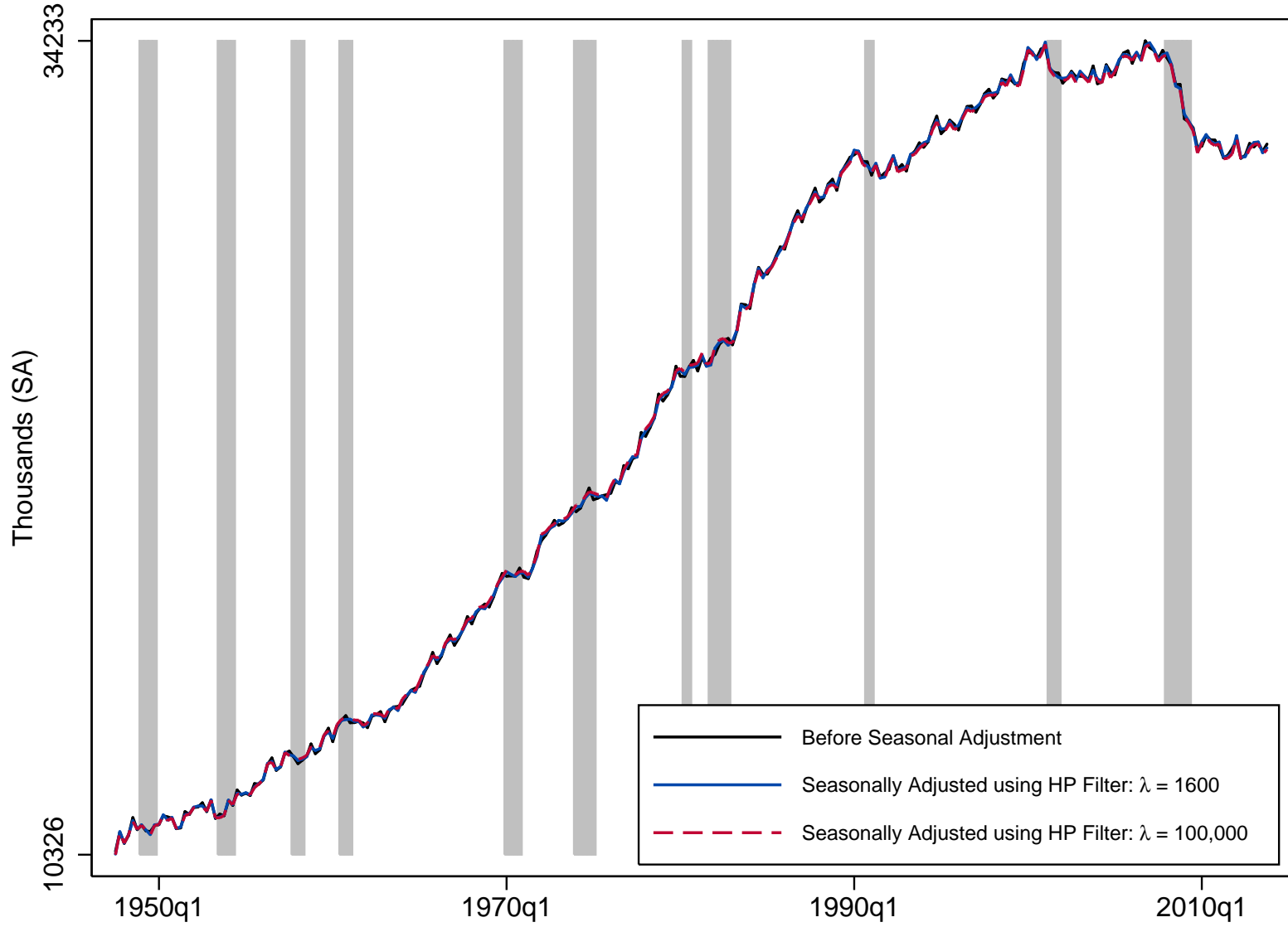


Figure A7. EMPLOYMENT OF ROUTINE COGNITIVE WORKERS (MIDDLE SKILL).

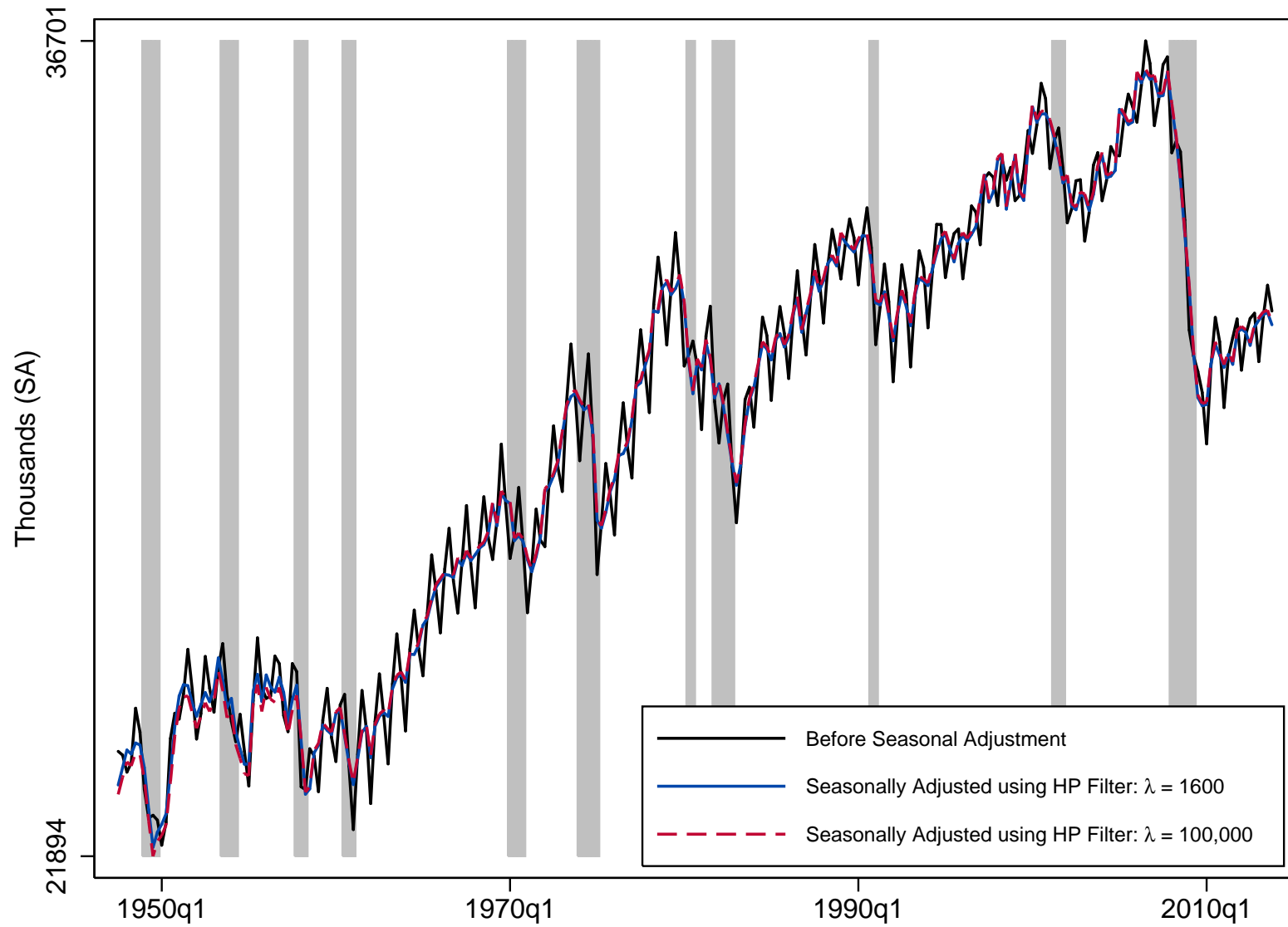


Figure A8. EMPLOYMENT OF ROUTINE MANUAL WORKERS (MIDDLE SKILL).

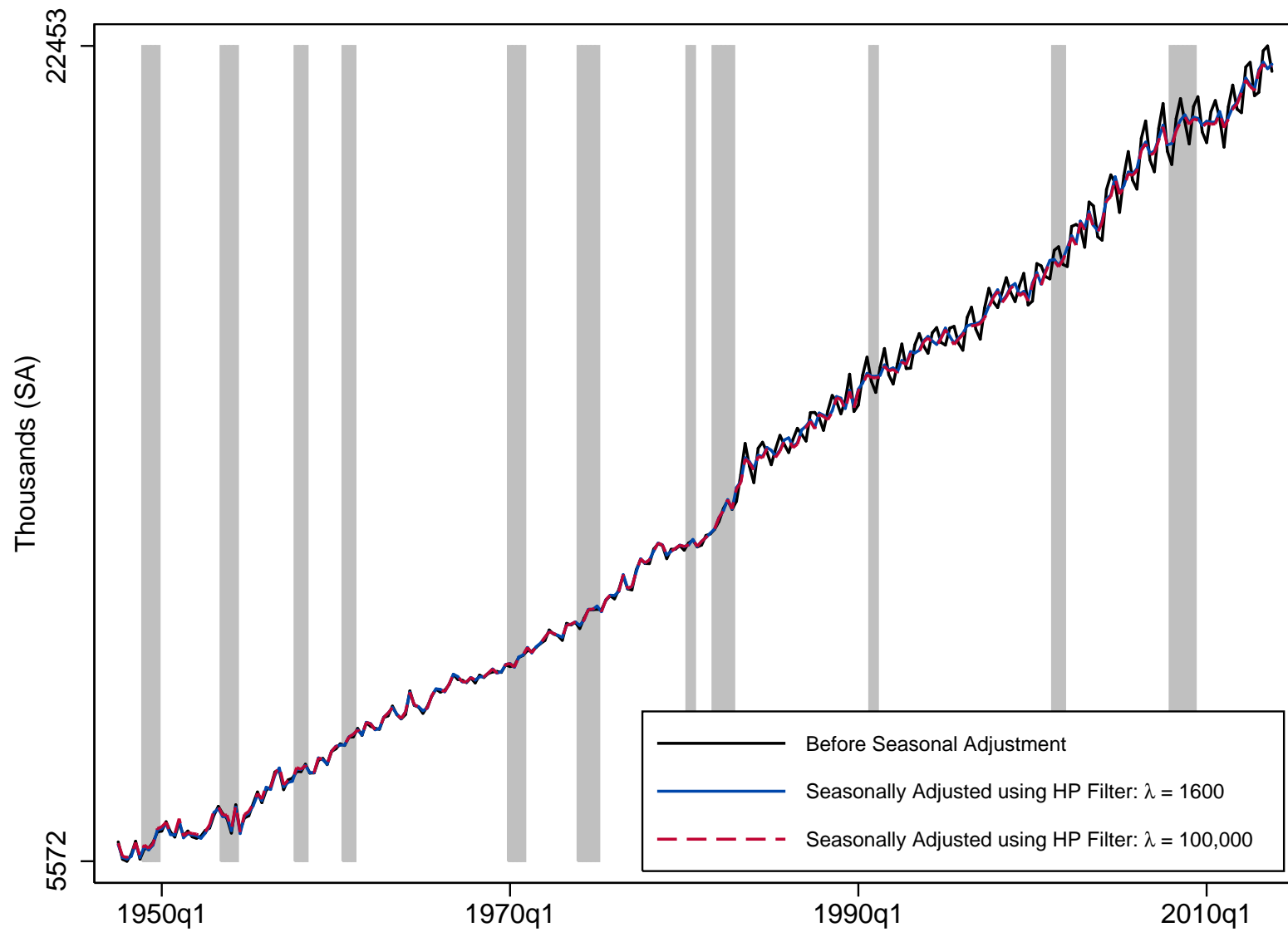


Figure A9. EMPLOYMENT OF NONROUTINE MANUAL WORKERS (LOW SKILL).

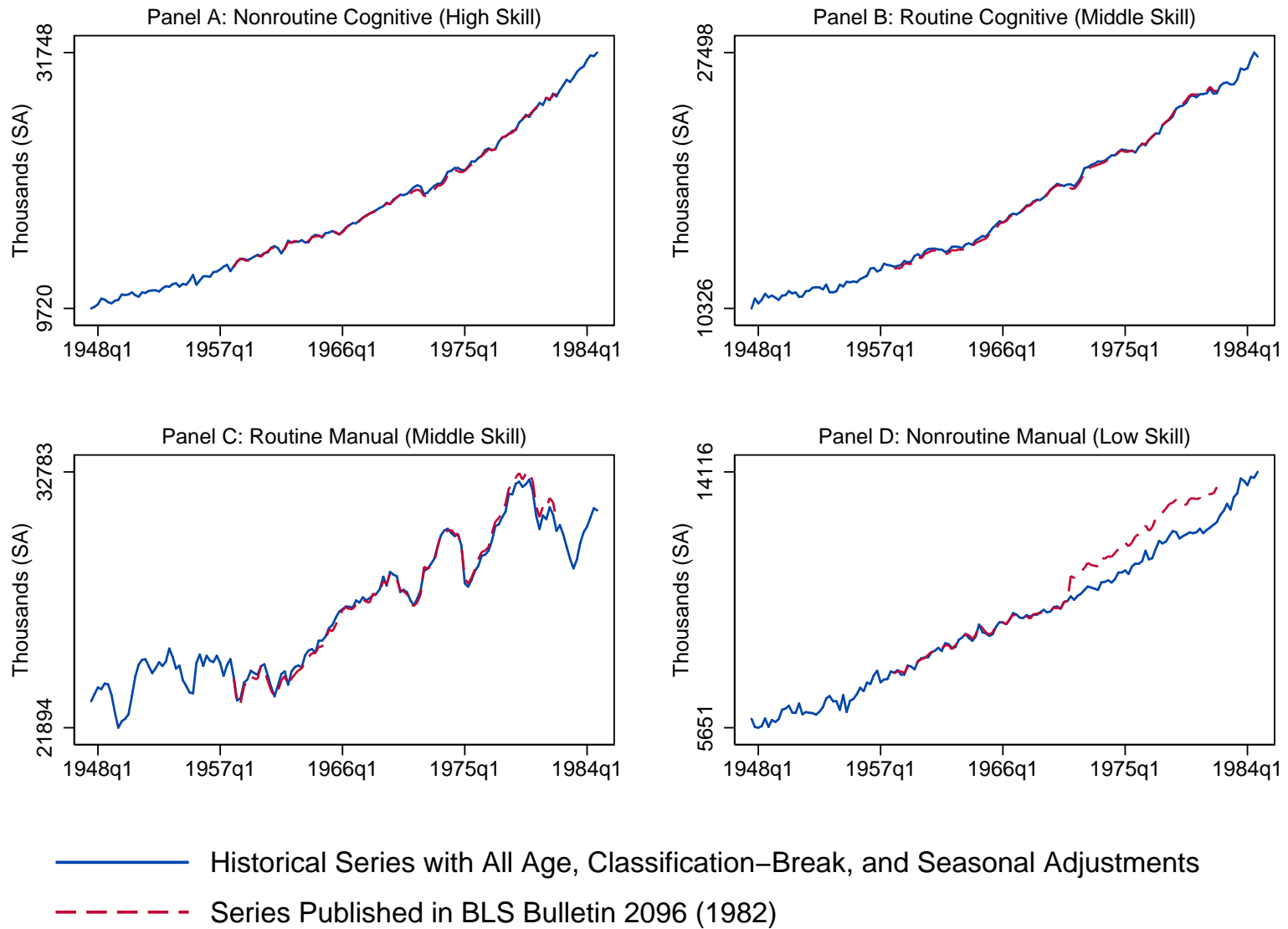


Figure A10. COMPARISON OF OCCUPATIONAL EMPLOYMENT SERIES WITH PUBLISHED SERIES FROM BLS BULLETIN 2096: 1958q1-1981q4. Note: All data are seasonally adjusted. Constructed employment series were seasonally adjusted using an HP filter with smoothing parameter $\lambda = 100,000$.

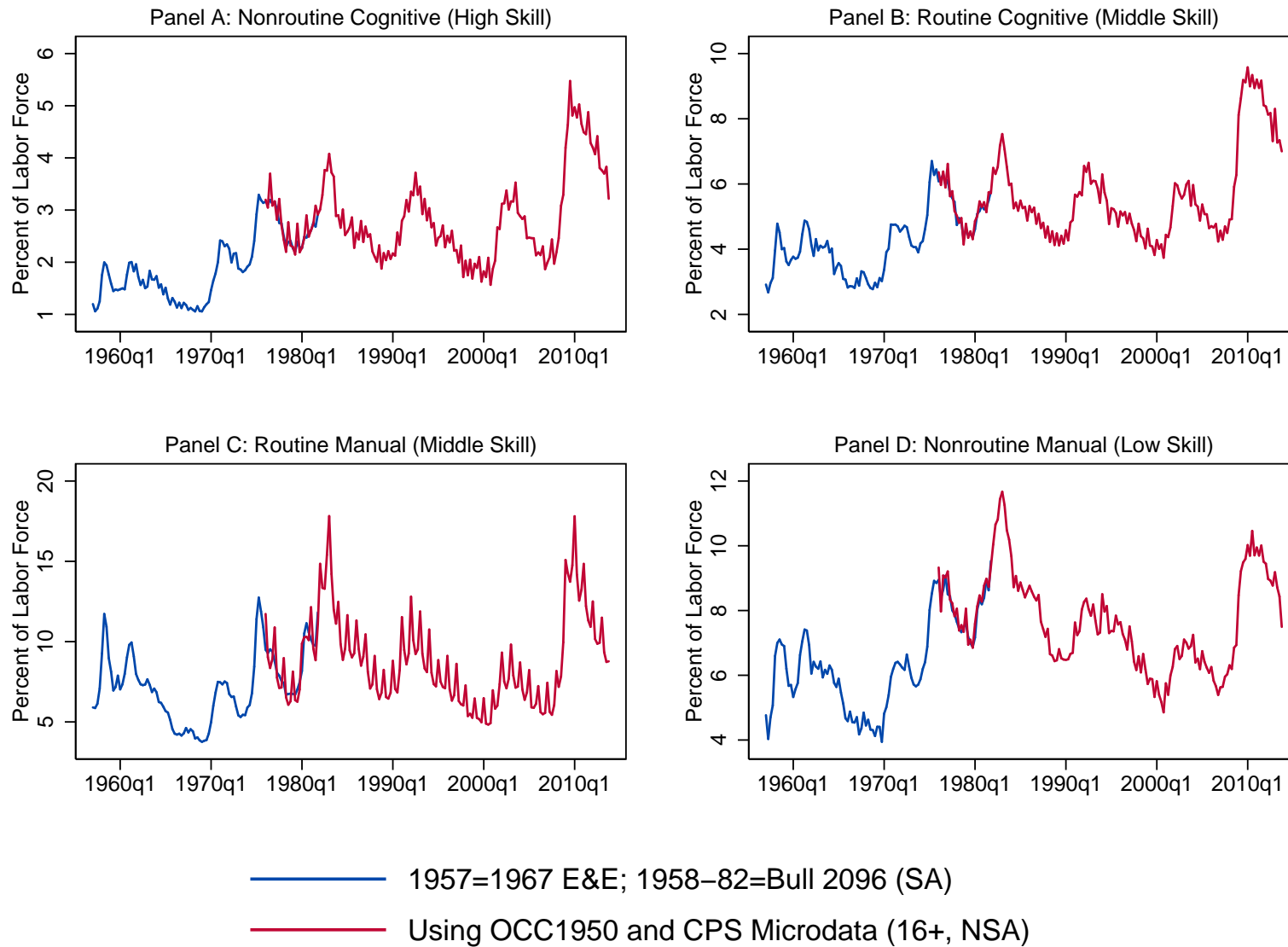


Figure A11. RAW DATA FOR UNEMPLOYMENT RATES: 1957q1-2013q3. Note: The source for the blue lines are Bulletin 2096, except for rates from 1957, which come from the March 1967 *Employment and Earnings*. The source of the red lines are CPS microdata using the OCC1950 coding system.

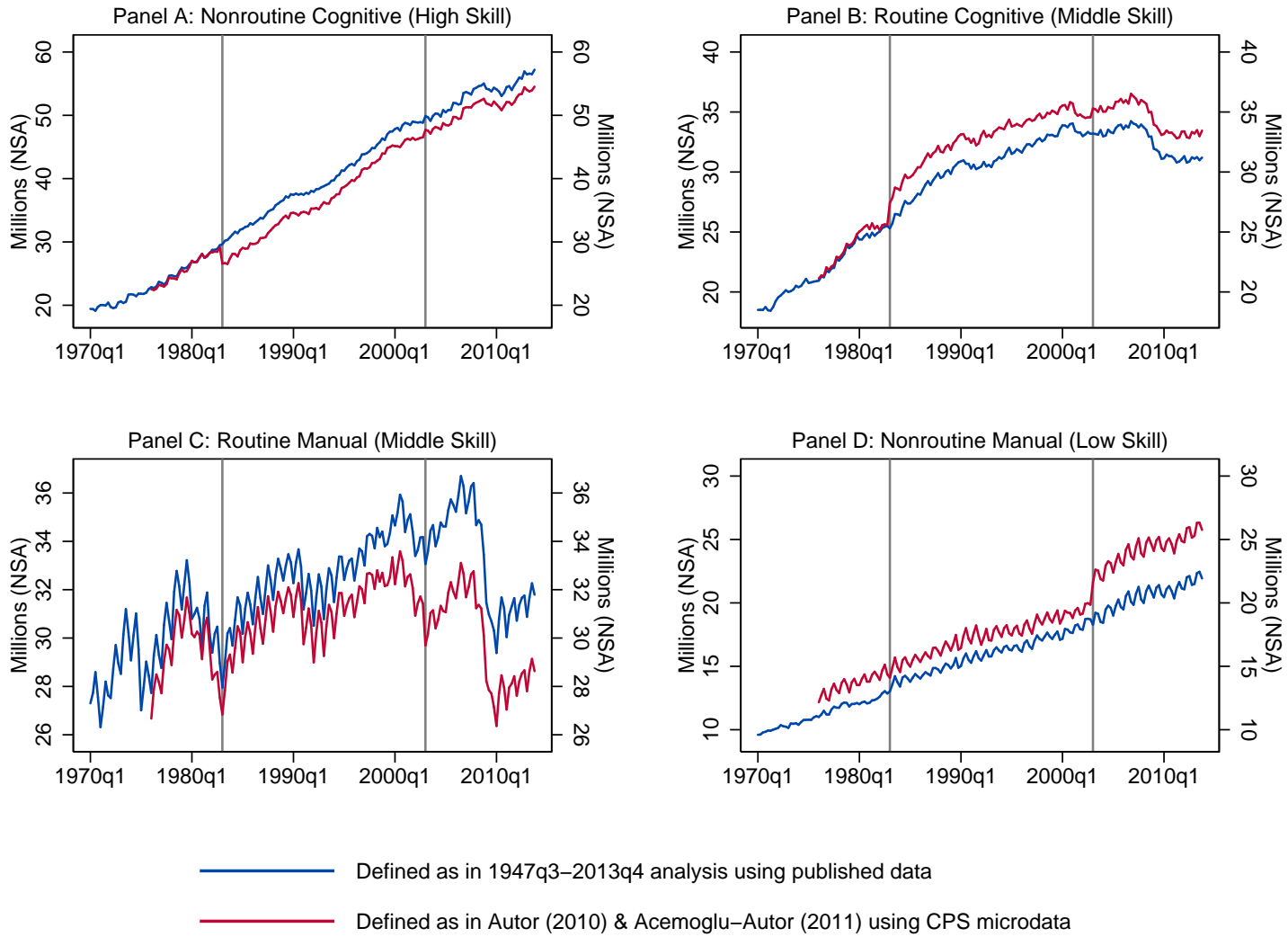


Figure A12. COMPARISON OF HISTORICAL DATA WITH CPS MICRODATA CODED AS IN AUTOR (2010) AND ACEMOGLU AND AUTOR (2011).