Business Cycle Properties of Job Polarization Using Consistent Occupational Data

Myungkyu Shim† and Hee-Seung Yang‡

Abstract
A significant obstacle to studying business cycle properties of job polarization has been the presence of inconsistencies in aggregate employment data for different occupation groups. In order to overcome this problem, we construct aggregate hours series using the method of 'conversion factors', which was originally developed by the Bureau of Labor Statistics. After showing that our data outperform previously available data in terms of consistency, we analyze two business cycle properties of job polarization that have not yet been studied before: (1) the changes in volatility of employment of each occupation group since the mid-1980s and (2) the asymmetric effects of recessions on employment of different occupation groups. We find that employment volatility of middle-skill occupations has decreased by 40% since the mid-1980s due to jobless recoveries observed in the last three recessions.

JEL classification: C82, E24, E32
Keywords: Business cycle; Job polarization; Consistency; Occupation; Conversion factor

† UC San Diego, email address: mkshim@ucsd.edu
‡ Monash University, email address: heeseung.yang@monash.edu

© 2014 Myungkyu Shim and Hee-Seung Yang
All rights reserved. No part of this paper may be reproduced in any form, or stored in a retrieval system, without the prior written permission of the author.
1 Introduction

One of the rapidly growing research areas in labor economics and macroeconomics is the study of heterogenous effects of labor market changes on workers who are employed in different occupations (jobs): high-skill occupations that require cognitive tasks, middle-skill occupations that require routine tasks, and low-skill occupations that require manual tasks.\(^1\) With this classification, Acemoglu and Autor (2011) and Autor (2010) find the emergence of ‘job polarization’ since the mid-1980s: jobs for the middle-skill occupation group have disappeared gradually, while those for the high-skill and low-skill occupation groups have consistently increased over time.

What then are the business cycle properties of job polarization? While long-run changes in the labor market have been studied extensively (See Autor, Katz, and Krueger (1998), Acemoglu and Autor (2011), Cortes (2012), Autor and Dorn (2013)), an analysis of how job polarization is connected to short-run fluctuations is rare: there are only two studies of which we are aware. Jaimovich and Siu (2013) show how job polarization is connected to jobless recoveries while Foote and Ryan (2012) analyze how the three occupation groups are affected differently by fluctuations by studying heterogenous responses of employment and job finding (separation) rates to business cycles. As noted by Foote and Ryan (2012), however, an inconsistent aggregate employment series makes it hard to analyze to what extent different occupation groups are disproportionately affected by economic downturns.\(^2\) This problem arises due to the fact that occupation classifications change over time (See Dorn (2009) and Foote and Ryan (2012)).

Our paper contributes to the literature on business cycle properties of job polarization in two ways. First, we provide consistent aggregate employment data for different occupation groups using the method of ‘conversion factors’. We present evidence that the constructed series of employment and total hours worked by occupation outperform previously available data in terms


\(^2\)In what follows, inconsistency of data refers to the situation where the data exhibit a sudden break when the occupation code changes.
of consistency. Second, using consistent occupational data, we provide information regarding the business cycle properties of job polarization that have not yet been documented. While a similar exercise is performed by Jaimovich and Siu (2013) and Foote and Ryan (2012), we are the first to document to what extent inconsistent aggregate employment data may fail to provide correct information about which occupations are more volatile than others at the business cycle frequency and which jobs are most affected by specific recessions.

We first address the consistency issue by comparing aggregate employment and total hours series for each of the three occupation groups mentioned above using the following two methods: (1) the ‘occ1990dd classification’ and (2) the ‘conversion factors’. The first method using the occ1990dd classification was originally suggested by Dorn (2009) and the second method using the conversion factors was originally developed by the Bureau of Labor Statistics (henceforth, BLS). We discuss these methodologies in detail in Section 2. For the purposes of comparison, we also construct aggregate employment series by occupation groups without applying either of these two methods, which we refer to as ‘raw’ data. By comparing aggregate hours variables obtained through each method, we find that only the method of conversion factors provides aggregate hours data without any break during the sample period. The occ1990dd data and the raw data, in contrast, exhibit breaks in aggregate data when the occupation code changes.

We then discuss the business cycle properties of job polarization. Previously available data did not allow researchers to study these properties because of problems of inconsistencies. In order to emphasize the role of consistent data, we compare the empirical findings from our data set with those from inconsistent data sets; in so doing, two particular aspects of business cycle fluctuations are studied in this paper: (1) the degree of changes in volatility of hours worked of each occupation group since the mid-1980s and (2) the asymmetric effects of recessions on employment of different occupation groups.

Suppose that we want to know if job polarization has made employment of a specific occupation group more or less stable at the business cycle frequency; i.e., a factor that causes job polarization at the long-run frequency may also cause each occupation group to be influenced by business cycles in a disproportionate way. This is equivalent to studying how volatility of
the hours variables has changed since the mid-1980s, which is in line with Castro and Coen-Pirani (2008) and Galí and van Rens (2011). We show, by using the Current Population Study (henceforth, CPS) Merged Outgoing Rotation Groups (henceforth, MORG) data, that measured changes in volatility are very inaccurate if inconsistent data are used. For instance, when the occ1990dd classification is used, employment volatility can be seen to have decreased the most for the high-skill occupation group since the mid-1980s. This implies that workers employed in this group benefit in two ways: the number of high-skill occupations has increased and, at the same time, such jobs have become more stable since the mid-1980s. When the conversion factor method is used to construct employment series, however, the decline in employment volatility for middle-skill occupations is the largest. We also discuss the implication of this finding for welfare costs of business cycles.

We then analyze the asymmetric effects of recessions on employment of different occupation groups, which could also be incorrectly measured with inconsistent data. As is emphasized by Jaimovich and Siu (2013), job polarization is connected to jobless recoveries and thus this information reveals another important business cycle property of job polarization. For instance, the middle-skill occupation group was affected the most by the 1980-81 recession when consistent data are used; on the contrary, the high-skill occupation group experienced the most volatile employment fluctuations during the 1980-81 recession when inconsistent data are instead used. Given that information regarding which occupation group hurts the most during a recession helps to identify the properties of the recession, understanding this discrepancy between different data sets informs studies of business cycles. We further show that the timing of the recent recessions coincides with changes in occupation codes, which raises again the importance of considering consistent data.

We do not argue that the conversion factor method is better than the occ1990dd classification system in every respect. As will be discussed in Section 2, we can apply the conversion factors only when we need ‘aggregate’ hours variables; i.e., it is not suitable for micro-level studies. Instead, the occ1990dd classification system, which constructs the balanced panel structure by
occupation, is more useful for micro-level studies, while aggregate hours variables constructed using the method exhibit breaks when occupation codes change. Hence, a researcher should be aware of the differences between the different methods and needs to choose the appropriate method that fits one’s research objective.

The paper is organized as follows. Section 2 describes in detail how we obtain aggregate hours variables for each occupation group. Section 3 shows how the data constructed by the conversion factors behave well from the perspective of consistency and Section 4 documents two important business cycle properties of job polarization. The final section is the conclusion.

2 Data Construction

In this paper, we consider three data sets of aggregate hours variables by occupation. The first data set is the raw data series to which no particular method is applied and this data set provides a benchmark in evaluating performance of different methods. The second data set is obtained using the occ1990dd classification that was suggested by Dorn (2009). The last data set, which we mainly use in this paper, is obtained using the conversion factors that was originally developed by the BLS. While the BLS also publishes aggregate employment data for different occupation groups which are constructed through the conversion factors on their website, there are two major shortcomings associated with using data directly from the BLS. First, its data set covers the period only after 1983. While one can find aggregate employment data for the periods before 1983, data for different periods are not directly comparable since the conversion factors are used only for data since 1983. Second, the BLS publishes employment data only. In this paper, we show that the conversion factors can be applied to the period before 1983 to construct consistent aggregate employment data and that the method can be also used to construct a total hours variable, which is not officially provided by the BLS.

The procedure we use to construct the aggregate hours variables is described in Appendix A.1.

---

3For example, Autor and Dorn (2013) use the occ1990dd classification to analyze longer-term changes in occupation shares based on micro data.
5Link: http://fraser.stlouisfed.org/publication/?pid=60.
In what follows, we first introduce data sources and then compare the occ1990dd classification system and the conversion factors in detail.

2.1 Data

There are three data sources used in this paper: the CPS MORG data\textsuperscript{6} that cover the period from Jan. 1979 to Dec. 2010; the CPS Basic Monthly data\textsuperscript{7} that cover the period from Jan. 1976 to Dec. 2010; and the (yearly) March CPS data that cover the period between 1971 and 2010. We restrict our data to 2010 since there was another major code change in occupational status in 2011. In addition, employment data from the March CPS cover the period 1971-2010 while the hours variable in the March CPS covers a shorter period, 1975-2009.\textsuperscript{8}

For the purposes of comparison of the different methods, we mainly use the CPS Basic Monthly data for employment (1976-2010) and the MORG data for total hours (1979-2010).\textsuperscript{9} In Appendix A.2, we provide additional figures obtained from the March CPS, which share the same properties reported in Section 3.

2.2 Comparison between the occ1990dd Classification System and the Conversion Factors

2.2.1 Overview

As is well-known, there have been several main changes in occupation codes in the Census: these have occurred in 1971, 1983, 1992, and 2003.\textsuperscript{10} With every introduction of classification systems, new detailed occupations were introduced, some of them were redefined, and some occupations were discontinued.\textsuperscript{11} These changes in occupational classification

---

\textsuperscript{6}Data were extracted from the NBER website: http://www.nber.org/data/morg.html.

\textsuperscript{7}Data were extracted from the NBER website: http://www.nber.org/data/cpsbasic.html.

\textsuperscript{8}Data were extracted from the IPUMS website: http://cps.ipums.org/cps, See King, Ruggles, Alexander, Flood, Genadek, Schroeder, Trampe, and Vick (2010).

\textsuperscript{9}The CPS Basic Monthly data contain information on labor force status but do not contain usual hours or wages except for the households in the outgoing rotation group.

\textsuperscript{10}The Census introduces a new classification of occupations every 10 years. The CPS uses the 1970 occupation codes from 1971 to 1982, the 1980 occupation codes from 1983 to 1991, the 1990 occupation codes for 1992 to 2002, and the 2000 occupation codes starting in 2003. For example, in January 2003, the CPS adopted the 2002 Census occupational classification system (the 2000 occupation codes) replacing the 1990 Census occupational classification system. Since the composition of detailed occupations changed substantially in the 2002 system compared with the 1990 system, direct comparison between the estimates from different classification schemes is not possible without major adjustments.

\textsuperscript{11}Even when the names of occupations appear to be the same, the definitions of the categories are sometimes different. For example, ‘Managers’ and ‘Farming Occupations’ were defined differently in 1990 from the way they were defined in 2000, even though these words appear in occupation tables from both Censuses (Scopp (2003)).
systems created complete breaks in employment series. As a result of the classification changes, occupational data with new occupation codes are not strictly comparable with earlier years.\textsuperscript{12}

In order to deal with this inconsistency problem, two methods can be used, although neither is perfect. One is to use the occ1990dd classification system created by David Dorn (See Dorn (2009)). The other is to use the conversion factors provided by the BLS, which are used to calculate the U.S. official employment data by occupation.

2.2.2 The occ1990dd Classification System The occ1990dd classification system provides a balanced panel of occupations covering the 1980, 1990, and the 2000 Censuses and the 2005 ACS.\textsuperscript{13} This system provides a mapping of Census occupation codes to a unified category system. That is, a single occupation code in a specific Census scheme corresponds to another single occupation code in a unified category system (occ1990dd).

The IPUMS-CPS website provides an occupation code ‘occ1990’, which is based on work by Meyer and Osborne (2005). Meyer and Osborne (2005) start from a weighted crosswalk (which is explained in the next section) and construct an unweighted (and thus easier to use) crosswalk that approximates the weighted correspondence. As Dorn (2009) points out, however, the occ1990 classification system does not provide a ‘balanced’ panel of occupations. For instance, ‘Economics Instructor’ was included in the occ1990 system, but there are no observations for this occupation in the 2000 Census because the specific fields of college teachers were not reported in the 2000 Census. Dorn (2009) shows that this unbalanced structure can be problematic for empirical work related to employment or wage changes within detailed occupations. Most of the improvements in the occ1990dd classification system were made through making occupation definitions ‘broader’ (by simply aggregating across occ1990 occupations) and providing more ‘consistent’ definitions in service occupations. That is, the occ1990dd system is primarily an aggregation of occ1990 codes in order to create time-consistent occupation categories and a balanced panel of occupations.

However, there is still a possible drawback associated with using the occ1990dd classification

\textsuperscript{12}While there was a significant change in the Census occupation coding scheme between the 1970 and 1980 Censuses and between the 1990 and 2000 Censuses, there were relatively few changes in the Census occupation coding scheme between the 1980 and 1990 Censuses, allowing for suitable comparisons over this period.

\textsuperscript{13}Some parts of discussions in this section are from Dorn (2009).
system. Since the occ1990dd code, which is a unified category system, provides a one-to-one correspondence between occupation titles in different Census schemes, it may not capture the fact that not every worker in a Census category falls into another single Census category. As shown in the next section, there exists a proportional flow between different Census categories. Therefore, the occ1990dd classification system may provide a biased estimate of labor market variables, especially when studying short-term fluctuations of aggregate occupational employment shares.

2.2.3 The Method of Conversion Factors  The conversion factors are weighted crosswalks which show the proportional flows for individual occupation categories between the two Census years in order to bridge changes in occupation codes. As illustrated by Blau and Liu (2013), this is important because all the incumbents of a particular Census occupation group do not necessarily match with the other Census occupation group, but rather they are split into several categories.

For a particular occupation within each Census, the Census crosswalks show lists of occupations in other Censuses. Since there is not always a one-to-one match between the Census classifications, we need to convert one scheme to the others. The BLS provides conversion factors to help data users bridge the gap created by breaks in occupational series. The conversion factors show the percentage distribution of employment from each occupation code in one classification, e.g., 1990, into each code in the other classification, e.g., 2000. BLS provides crosswalk tables and conversion factors between the 1990 and 2000 Census classifications. These factors are based on three-year average survey microdata (2000-2002, double-coded sample) that were coded to both the old and new classification systems. This process puts each person in the sample into both classification sets. The conversion factors, thereby, provide information on the proportion of actual workers that went from one Census category into another.\textsuperscript{14}

For example, among ‘Executive, Administrative, and Managerial’ workers in the 1990 oc-

\textsuperscript{14}The BLS uses a similar method to create conversion factors for earlier Censuses with the 1970 Census Sample comprising 127,125 individuals drawn from the 1970 Census (Mosbacher and Ortner (1989)). This sample was used to determine what fraction of the cases with a given 1970 code corresponds to each of several 1980 codes. For converting 1980 to 1990 classifications, only a few changes were needed to make the 1980 codes compatible with the 1990 classifications as there were relatively few changes in the Census occupational coding scheme between the 1980 and 1990 Censuses.
ocpiration group, the Census crosswalk distributes 73.1 percent of incumbents to ‘Management, Business, and Financial Operations Occupations’ in the 2002 occupational group, 11.1 percent to ‘Office and Administrative Support Occupations’, 4.4 percent to ‘Sales and Related Occupations’, and 4.1 percent to ‘Professional and Related Occupations’. These percentages could be applied to historical employment data with the 1990 classification to approximate employment with the 2002 classification.\(^{15}\)

In theory, it is also possible to convert the 2000 categories into those for 1990 in order to make reverse comparisons over time.\(^{16}\) However, the Census Bureau recommends using the Census crosswalk only for converting the codes forward to display employment of 1990 in 2000, as the 2000 Census is more up-to-date.

Although the conversion factors provide some linkages between the old and new classifications, there are undoubtedly some limitations in how they can be used. Scopp (2003) reports that the conversion factors are subject to sampling errors, especially when the numbers for a detailed category are very small. Thus, the resulting series is only an approximation providing a general employment trend over time.\(^{17}\) Also, the double-coding process might involve coding errors. These errors may contaminate comparisons across classifications. In addition, while the conversion factors allow us to analyze employment and total hours worked by detailed occupations, this method is not appropriate for other variables such as average hours worked and real wages by occupation. For instance, if we apply the conversion factors to construct the hourly wage rate by each occupation group, we can observe breaks between 2002 and 2003, when there was a change in occupation codes. Nevertheless, the conversion factor method has a few advantages compared to other methods, which will be shown in the following section.\(^{18}\)

\(^{15}\)The BLS also created conversion factors for industry employment. For example, for the construction industry, 92 percent of employment in the old construction classification was re-classified to the new construction classification. The remaining 8 percent of employment was re-classified among other industry categories in the new classification (Link: http://www.bls.gov/cps/constio198399.htm).

\(^{16}\)See Scopp (2003), p.5.

\(^{17}\)Link: http://www.bls.gov/cps/cpsoccind.htm.

\(^{18}\)The reason that crosswalks, like conversion factors, are necessary is that the Census has changed the occupation codes every 10 years, and the conversion factors are developed to make a consistent series of occupations that span longer periods. If researchers are interested in doing an analysis for a short time period in which there is one consistent set of Census occupation codes (e.g., 1992 - 1999), then they should use the original code, occ1990, provided by the IPUMS, not occ1990dd or conversion factors. If their analysis spans multiple time
3 Comparison: Which Method Yields More Consistent Aggregate Data?

In this section, we present figures for the aggregate hours variables (employment and total hours) to show that aggregate data constructed by the conversion factors are more consistent than other data sets.\textsuperscript{19}

3.1 Employment We first consider employment, which uses the CPS Basic Monthly data that covers more periods than does the MORG data. See Figures 3.1 to 3.3. In each figure, the solid blue line is the employment series constructed using the conversion factors, the thick green dotted line is constructed using the occ1990dd classification, and the thin red dotted line is from the raw data. The shaded regions are the official NBER recession dates.

![Figure 3.1: Employment: High-Skill Occupations](image)

We can observe that the employment series constructed using the occ1990dd classification periods, however, they have to work with a relatively consistent crosswalk.

\textsuperscript{19}To download consistent aggregate data, visit the authors’ website.
Figure 3.2: Employment: Middle-Skill Occupations

Figure 3.3: Employment: Low-Skill Occupations
have four breaks over time. First and second, the employment of high-skill occupations exhibits one big break between 1982 and 1983 and one (relatively) small break between 2002 and 2003. Two other breaks are observed in Figure 3.3: the employment of low-skill occupations exhibits one small break between 1992 and 1993 and one big break between 2002 and 2003. In addition, we observe that the level of employment for high-skill occupations is lower for data constructed using the occ1990dd classification than for data constructed using the conversion factors since 1983, while the level of employment for low-skill occupations constructed using the occ1990dd classification is lower than the other series constructed using the conversion factors until 2002.

Similarly, the raw data exhibit two breaks; one for high-skill occupations between 1982 and 1983 and the other for low-skill occupations between 2002 and 2003. This implies that the occ1990dd classification system does not resolve the inconsistency problem associated with using aggregate employment data; rather, it amplifies the inconsistency problem. In contrast to these data sets, the data we construct using the conversion factors exhibit no breaks in any of the series; i.e., the conversion factor method yields the most consistent employment series among the three methods. Applications of the above findings, which are presented in Section 4, will make it clear why having a more consistent data set is important in studies of business cycle properties of job polarization.

3.2 Total Hours Worked Now we move our attention to total hours worked for each occupation group. In what follows, we use the MORG data instead of the CPS Basic Monthly data because the latter data set does not provide information on hours worked. Figures 3.4 to 3.6 are the corresponding figures.

From Figure 3.4, we can observe that total hours worked for the high-skill occupation group constructed using the conversion factors (thick blue line) has a very similar pattern to that constructed using the occ1990dd classification (dotted green line). However, our data do not exhibit any breaks, while the dotted green line shows two major breaks: one during 1982-1983 and the other during 2002-2003. Given that other total hours series for the middle-skill and low-skill occupation groups are almost identical between the two methodologies (Figures 3.5 and 3.6), this fact confirms that the conversion factor method yields a more consistent total hours
worked series than does the occ1990dd classification system. Note further that the raw data (dotted red line) sometime outperform the occ1990dd data from the perspective of consistency.

**Figure 3.4: Total Hours: High-Skill Occupations**

**Figure 3.5: Total Hours: Middle-Skill Occupations**
In this section, we study two business cycle properties of job polarization using the consistent data set. Consider first a researcher who is interested in long-run changes (trends) in the labor market variables. Then, the breaks we observe from the pre-existing data may not be an issue since they do not alter conclusions one can obtain from inconsistent data. For instance, if a researcher is only interested in the occurrence of job polarization, the disappearance of middle-skill occupations during the last 30 years is observed whether data are consistent or not. As we show below, however, observations on business cycle properties of job polarization change dramatically when we use consistent occupational data instead of using inconsistent data. This inconsistency problem has been an obstacle to analyzing business cycle properties of job polarization.

Hence, the findings reported in this section do not just reveal the importance of having consistent data, but they do deepen our understanding about how long-run changes in the labor market (i.e., job polarization) are related to business cycles. In so doing, we first document the

---

20The growth rate of employment may differ across different data sets, but it does not hide job polarization.
changes in volatility of employment of each occupation group since the mid-1980s.\textsuperscript{21} Then, we study the extent of the heterogeneous effects of recessions on different occupations. In particular, we find that the timing of the changes in occupation codes coincides with the three recent recessionary episodes so that using inconsistent hours variables leads to misunderstanding about the effects of recessions on the different occupation groups.

4.1 Changes in Volatility of Occupational Employment In this section, we analyze which jobs are less volatile than others at the business cycle frequency and whether there has been any change in that information since the mid-1980s. The first question is translated into the problem of seeking jobs that fluctuate less than others during business cycles. The second question is noteworthy for the following reasons. Job polarization, which shifts firms’ demand from the middle-skill group to the high-skill and low-skill groups, started to occur since the mid-1980s and it is possible that such job polarization also affects the volatility of jobs in an asymmetric way. For instance, the middle-skill occupation group, whose importance in production decreases over time, may suffer more from business cycle fluctuations than before, while other groups do not. Hence, knowledge of the changing cyclical properties of employment deepens our understanding of job polarization.\textsuperscript{22} Moreover, given that changes in the aggregate economy since the mid-1980s such as the great moderation and labor market changes - for instance, the vanishing procyclicality of labor productivity (see Galí and van Rens (2011)) and the increase in wage volatility relative to GDP (see Champagne and Kurmann (2013)) - are interesting to researchers, a study of changing cyclical properties of jobs (if any) divided by occupation further increases our understanding of the labor market.

For the exercise, we use the employment series obtained from the MORC CPS, from 1979 to 2010. We first remove the seasonality of data by X-12 ARIMA and then detrend the data by using the Baxter and King filter with $\kappa = 12$ (size of the window for the filter). We set the lowest frequency as 6 quarters and the highest frequency as 32 quarters as usual.\textsuperscript{23} We then compute the standard deviation of each detrended employment series by dividing the whole period into

\textsuperscript{21}A similar conclusion can be drawn if we use the total hours worked.

\textsuperscript{22}In this sense, the application in this subsection is related to Jaimovich and Siu (2013).

\textsuperscript{23}Results do not change even when we use the Hodrick-Prescott filter.
two periods: (1) 1979-1983 and (2) 1984-2010, following Castro and Coen-Pirani (2008) and Gali and van Rens (2011). One might raise a concern about the validity of the statistics from the first subperiod because it only covers five years. As a robustness check for our results, we do the same exercise with the March CPS of which employment series cover the period from 1971 to 2010 so that the first subperiod is much longer; the result is reported in Table A.1.\footnote{For the March CPS we set 2 as the lowest and 8 as the highest frequency with $\kappa = 3$.} We find that the results from the two data sets are almost identical.

Table 4.1 shows the main results.

<table>
<thead>
<tr>
<th>Occupation Group</th>
<th>High-Skill</th>
<th>Middle-Skill</th>
<th>Low-Skill</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methodology\Period</td>
<td>(1) 79-83</td>
<td>(2) 84-10</td>
<td>(1) 79-83</td>
</tr>
<tr>
<td>Conversion Factors</td>
<td>0.39</td>
<td>0.33 (0.83)</td>
<td>0.84</td>
</tr>
<tr>
<td>occ1990dd System</td>
<td>1.09</td>
<td>0.48 (0.45)</td>
<td>0.88</td>
</tr>
<tr>
<td>Raw Data</td>
<td>1.11</td>
<td>0.37 (0.44)</td>
<td>0.87</td>
</tr>
</tbody>
</table>

Note: 1. All numbers are multiplied by 100.
2. Numbers in parentheses are the ratios between (2) and (1).

Stylized Fact 1 summarizes the key observation from the first row, where the volatility of consistent employment series is reported.

**Stylized Fact 1** (Volatility Changes: Method of Conversion Factors). Let $\Delta \sigma_i$ be the percentage change in the employment volatility of occupation group $i$. Then $\Delta \sigma_M < \Delta \sigma_H < 0 < \Delta \sigma_L$; i.e., the change in volatility of detrended employment since the mid-1980s is the lowest for the middle-skill occupation group.

A key finding from the consistent data is that employment volatility decreases the most for the middle-skill occupations. In this section, we focus on the implication of the above finding for welfare costs of business cycles.\footnote{This part is closely related to our companion paper, Shim and Yang (2013). In this paper, we argue that the above finding can explain the ‘non-monotonic’ changes in employment volatility by skill (education) group since the mid-1980s.} Note that lower employment volatility is usually interpreted as ‘good’ in the macroeconomics literature when the asset market is incomplete since workers...
are risk-averse and lower labor income volatility implies less fluctuations in consumption.\textsuperscript{26} In this sense, the observed change in employment volatility since the mid-1980s seems to be good news for workers who are employed in middle-skill occupations: the employment volatility for middle-skill occupations declines by 40 percent, while employment volatility decreases slightly for high-skill occupations and changes little for low-skill occupations.

As emphasized by Shim and Yang (2013), however, the large decline may not be good news for middle-skilled workers when the business cycle property of job polarization is taken into account. Figure 4.1 shows the main finding of Jaimovich and Siu (2013) and this is the key to understanding our findings. In this figure, we divide the employment of each occupational group using the MORG data by civilian noninstitutional population,\textsuperscript{27} and then filter the series with the Baxter-King filter and use the trend components. The shaded regions in the figure indicate the NBER official recession dates.

![Figure 4.1: Employment to Population: Middle-Skill Occupations](image)

During the recovery, after the early recession (the 1981-82 recession), employment to pop-

\textsuperscript{26}Here, we assume that labor supply is inelastic. In addition, we postulate that wage premium (skill premium) is acyclical for this statement to be true as evidenced by Castro and Coen-Pirani (2008). See Lucas (1987), Mukoyama and Sahin (2006), Castro and Coen-Pirani (2008), and Shim and Yang (2013) for detailed discussion.

\textsuperscript{27}Downloaded from the BLS website. We abstract high- and low-skill occupations; one can find the figures for the two occupation groups in Shim and Yang (2013) or Jaimovich and Siu (2013).
ulation ratio of the middle-skill group goes back to its original level. During the last three recessions (the 1990-91, 2001, and 2007-2009 recessions), however, once the negative shock hits the economy, employment (to population) decreases and never comes back to its original level. In other words, the losses of middle-skill jobs (occupations) are concentrated during recessions, suggesting that, consistent with Jaimovich and Siu (2013), job polarization is a business cycle phenomenon. This is the main driver of the above finding; during the recovery episode without a jobless recovery, the usual filtering method detects the quick rebound of employment as a big swing of the detrended series. During the jobless recovery episode, however, employment stays at the new lower level, and hence, the changes in the detrended series are much more nuanced. As a result, the standard deviation of detrended employment is computed to be much smaller in the economy with a jobless recovery.

Therefore, the ‘lower’ employment volatility of middle-skill occupations may not be a welfare-improving change. As jobs do not come back after the recession, the workers who used to work in middle-skill occupations need to find other jobs in either high- or low-skill occupations and it takes time to learn new skills to perform tasks different from the previous tasks they did. Hence, if the unemployment period of these workers becomes longer, the welfare cost of business cycles may become larger instead of becoming smaller.\footnote{The welfare cost of business cycles is larger for long-term unemployed workers. See Krusell, Mukoyama, Şahin, and Smith (2009) for related discussions.}

What then do we learn if we instead use inconsistent data? It is easy to see from Table 4.1 that one will draw very different conclusions from different data sets. Stylized Fact 2 summarizes the result.

**Stylized Fact 2** (Volatility Changes: occ1990dd Classification and Raw Data). $\Delta \sigma_H < \Delta \sigma_M < 0 < \Delta \sigma_L$ when inconsistent data are used; i.e., employment volatility decreases significantly for high-skill and middle-skill occupations while it increases significantly for low-skill occupations.

When employment data are constructed using the occ1990dd classification or the raw data are used, the decline in the standard deviations of employment is larger for high-skill occupations than for middle-skill occupations (the standard deviation for high-skill occupations decreases to
less than half of the previous level). Notice that the consistent data set shows smaller decline (the volatility of employment for high-skill occupations decreases by less than 20 percent). The differences are also dramatic when the low-skill occupation group is considered. Employment data constructed using the conversion factors show that there has been virtually no change in volatility for the low-skill group. The volatility for the low-skill group, however, increases by about 30 percent for the occ1990dd data and for the raw data. In contrast to these two groups, there are no big differences for the middle-skill occupation group.

In order to emphasize the importance of knowing the information discrepancies, consider a researcher who uses the occ1990dd data or the raw data. Then, her conclusion from Table 4.1 will be that (1) employment of the high-skill occupations has become dramatically stable in the second subperiod, while (2) that of the low-skill occupations has become much more volatile. Notice that job polarization is the phenomenon that employment is increasing for both high-skill and low-skill occupations. Hence, the observations in Table 4.1, when the occ1990dd data or the raw data are used, lead the researcher to conclude that there exist some factors that make the employment of the high-skill occupation group less volatile, whereas they make the employment of the low-skill group more volatile at the business cycle frequency, even though the labor demand increases for both of them over time. Notice that the long-run labor market changes are favorable for low-skill occupations; both wages and available jobs have increased for those occupations. Contrary to these favorable long-run changes, the observations from inconsistent data imply that the changes in the short-run property of the labor market have affected low-skilled workers negatively.

These findings, however, are no longer valid if we use a consistent employment series that is constructed using the conversion factors. Contrary to the above implications, the consistent data show that there has not been a dramatic change in employment volatility for both high-skill and low-skill occupations; i.e., changes in the labor market at the business cycle frequency affect these two occupation groups in a similar way.

Figure 4.2 shows the reason that we draw such different conclusions from the different data sets. Recall that the employment series constructed using the occ1990dd classification and raw
data exhibit breaks between 1982 and 1983 and between 2002 and 2003, as in Figures 3.1 and 3.3. These are recognized as a big recession or expansion by the detrending method (Baxter-King filter) so that we observe big fluctuations between 1982 and 1983, as one can easily observe in Figures 4.2a and 4.2b. This is the reason that the volatility changes are similar for the middle-skill occupation group among the three methods: none of the three method exhibits breaks for the middle-skill occupation group (See Figure 3.2). Figure 4.2b also explains why the standard deviation of employment for the low-skill occupations increases by about 30 percent when inconsistent data are used: there is a large swing between 2002 and 2003, which increases the volatility, and this is the consequence of the artificial break in data when they are constructed using the occ1990dd classification or constructed without any particular method.

![Figure 4.2a: Detrended Employment, High-Skill Group](image1)

![Figure 4.2b: Detrended Employment, Low-Skill Group](image2)

![Figure 4.2: Detrended Employment Series](image3)

Table 4.2 illustrates the above discussion more clearly. In order to study the extent of the effects of breaks on standard deviations, we change the span of the first subperiods: 1979-1981, 1979-1982, and 1979-1983 (the benchmark case). The numbers in Table 4.2 confirm our discussion. The volatilities of employment for the high-skill occupation group, which are constructed using occ1990dd system or by no method (raw data), decline dramatically as we shorten the first period to remove the effect of the breaks between 1982 and 1983. Notice that our data set constructed using the conversion factors preserves the similar magnitude of the standard
deviations across the different periods, which comes from the fact that our data are consistent regardless of the changes in the coding scheme.

Table 4.2: Standard Deviations of the Detrended Employment: Different Base Years

<table>
<thead>
<tr>
<th>Methodology</th>
<th>79-81</th>
<th>79-82</th>
<th>79-83</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>H</td>
<td>M</td>
<td>L</td>
</tr>
<tr>
<td>Conversion Factors</td>
<td>0.29</td>
<td>0.71</td>
<td>0.33</td>
</tr>
<tr>
<td>occ1990dd System</td>
<td>0.41</td>
<td>0.69</td>
<td>0.33</td>
</tr>
<tr>
<td>Raw Data</td>
<td>0.46</td>
<td>0.76</td>
<td>0.40</td>
</tr>
</tbody>
</table>

Note: 1. All numbers are multiplied by 100.
2. H stands for high-skill, M stands for middle-skill, and L stands for the low-skill occupation group, respectively.

4.2 Asymmetric Effects of Recessions on Different Occupation Groups

In this section, we consider another important aspect of business cycles: heterogenous effects of recessions on different jobs. This is an important question in defining characteristics of a recession. For example, the 2007 recession was originated from the collapse of the housing market, hence, the construction and financial sectors were mostly affected by the recession. Meanwhile, the health care sector expanded during the recession (Sahin, Song, and Hobijn (2010)). In summary, as Jaimovich and Siu (2013) have already found, the 2007 recession affected each occupation disproportionately and deepened job polarization.

However, the previous inconsistent data sets did not allow careful analysis in this regard: recall that the major code changes in occupations occurred in 1983, 1992, and 2003. Unfortunately, these changes in occupation codes nearly perfectly coincide with the three recessions whose NBER official dates are July 1981-November 1982, July 1990-March 1991, and March 2001-November 2001, respectively. Hence, inconsistent data with breaks in these periods will provide us with inaccurate information on which group was mostly affected by these recessions. Our consistent data, however, makes it possible to analyze the asymmetric effect of recessions.

The increase in the standard deviation of employment for the middle-skill occupation group, which is the common feature across the three data sets, is the consequence of the 1981-82 recession, which mainly affected the jobs that were included in the middle-skill groups.
For this exercise, we use the detrended employment series but focus on these specific episodes: the 1981-82 recession, the 1990-91 recession, and the 2001 recession. Figures 4.3 to 4.5 show the results. In this exercise, we compare the employment series constructed by using the conversion factors with the series obtained by applying the occ1990dd classification. The shaded regions are the official NBER recession dates, the solid blue line is the employment of the high-skill occupation group, the thick dotted green line is the employment of the middle-skill occupation group, and the thin dotted red line is the employment of the low-skill occupation group.

We first consider the 1981-82 recession. Figure 4.3a shows that the group that loses jobs the most is the middle-skill occupation group, while the high-skill occupation group is not significantly affected by the recession. Figure 4.3b, however, provides completely different information about labor market fluctuations: the employment fluctuations of the high-skill occupations are greater than those of the middle-skill occupations. Given that the 1981-82 recession was asymmetric across sectors and the manufacturing sector, which depends heavily on the middle-skill occupation group, was affected most by the recession, we can conclude that Figure 4.3b gives

---

30 We use the CPS Basic Monthly data in this section.
31 We get similar results when we use the total hours series.
us inaccurate information about the differences in employment fluctuations across occupation groups during the 1981-82 recession. This is the natural consequence of the break in data constructed using the occ1990dd classification, as is evident in Figure 4.2a.

How about the 1990-91 recession? Similar to the 1981-82 recession, it is evident that different information is drawn from different data sets. Recall that there was a break in employment of low-skill occupations in Figure 4.2b; because of this break, the negative effect of the recession on low-skill occupations is exaggerated when employment data are constructed using the occ1990dd classification. Figure 4.4b shows that employment of low-skill occupations declines by about 2 percent after the end of the recession. This is slightly greater than the decline in employment for middle-skill occupations (less than 2 percent); hence, it may seem that the 1990-91 recession affected low-skill occupations the most in terms of employment fluctuations. It is the case, however, that employment for the low-skill occupation group declines by about 1 percent when consistent employment data are instead used, which is smaller than that of middle-skill occupations which experienced about a 1.3 percent decline in employment.

Figure 4.5 plots employment fluctuations of each occupation group for another recessionary episode, the 2001 recession. One can easily observe that the two graphs in the figures show different patterns of recovery from the recession. Figure 4.5a, derived from consistent data, shows that employment of the low-skill occupation group recovers by 2 percent from the trough to
the top during the recovery, while employment of the high-skill occupation group changes little. However, Figure 4.5b, constructed using the occ1990dd classification, shows that the employment of the low-skill occupation group recovers by more than 4 percent during the recovery, which is double what we observe from the consistent data. Furthermore, if we use the occ1990dd data without considering the break between 2002 and 2003, we can observe that the high-skill occupation group exhibits more than a 2 percent drop in employment during the recovery, suggesting the misleading conclusion that there was a lagged response of the high-skill occupation group to the recession. However, this is just a reflection of the artificial break, which we find in Figures 4.2a and 4.2b.

Notice that the above analysis using ‘detrended’ data is not conducted by Foote and Ryan (2012) because of the problem of data inconsistency. Hence, discussions in this section supplement their analysis.

5 Conclusion

We have discussed the value of having consistent aggregate employment data through the method of conversion factors in studies of business cycle properties of job polarization. In terms of methodology, we show that: (1) the conversion factor method can be also applied to construct
total hours worked; (2) it can be applied to the period before 1983 where such data are not provided by the BLS; and (3) employment and total hours worked, which are constructed using the conversion factors, are more consistent than previously available data sets. Equipped with consistent data, we study two aspects of job polarization over the business cycles, which improve our understanding of job polarization. In particular, we investigate (1) the changes in volatility of hours worked of each occupation group since the mid-1980s and (2) the asymmetric effects of recessions on employment of different occupation groups. The results show that while the vast majority of job losses were in the middle-skill occupation group during most recessions, their employment volatility has decreased a lot since the mid-1980s due to jobless recoveries observed in the last three recessions.

Notice that while the occ1990dd classification system is not successful in constructing consistent aggregate data, it is more useful in micro-level studies that require unambiguous assignment of individual workers to specific occupation codes. By contrast, the conversion factor method cannot be used in micro-level studies since it is only useful for constructing aggregate data instead of tracking each individual. Hence, a researcher should be cognizant of the different advantages of using different methodologies when analyzing job polarization.
REFERENCES


25


A Appendix

A.1 Aggregate Data Construction  The March CPS, the MORG data, and the CPS Basic Monthly data report a person’s primary occupation. Respondents who held more than one job were required to report the job in which they worked the largest number of hours. Following Autor (2010), we assign ‘Executive, Administrative, and Managerial Occupations’, ‘Professional Specialty Occupations’, and ‘Technicians and Related Support Occupations’ to the high-skill occupation group; ‘Sales Occupations’, ‘Administrative Support Occupations, including Clerical’, ‘Precision Production, Craft, and Repair Occupations’, ‘Machine Operators, Assemblers and Inspectors’, ‘Transportation and Material Moving Occupations’, and ‘Handlers, Equipment Cleaners, Helpers and Laborers’ to the middle-skill occupation group; and ‘Private Household Occupations’, ‘Protective Service Occupations’, and ‘Service Occupations except Protective and Household’ to the low-skill occupation group. We exclude ‘Farming, Forestry, and Fishing Occupations’ and ‘Military Occupations’.33

The main variables of interest are constructed as follows:

1. Employment: In the CPS, individuals’ employment status was determined on the basis of answers to a series of questions relating to their activities during the preceding week. Those who reported doing any work at all for pay or profit are classified as ‘employed’. We aggregate employment by their occupations in a given month or year with their sampling weight:

\[
Employment_{D,t} = \sum_{i \in D} 1_{\text{employed},i} u_{i,t} \quad (A.1)
\]

where \(1_{\text{employed},i}\) indicates individuals’ employment status, which equals one when the individual \(i\) is employed at time \(t\) and zero when he or she is unemployed. \(D\) is the individual’s

---

33The categorization of workers in Cortes (2012) is almost identical to ours, but the author uses the terms ‘non-routine cognitive (high-skill occupation)’, ‘routine (middle-skill occupation)’, and ‘non-routine manual(low-skill occupation)’.
group category by occupation. $u_{i,t}$ is the individual sample weight.\(^{34}\)

2. **Total hours worked**: We compute the total hours of work using the earnings weight provided by the CPS as follows.

$$
TotalHours_{D,t} = \sum_{i \in D} h_{i,t} u_{i,t}
$$

(A.2)

where $h_{i,t}$ is weekly hours worked for individual $i$ at time $t$.

### A.2 Additional Figures and Table from Other Data Sources

We provide figures for employment and total hours worked from the March CPS in Figure A.1 to A.3. One can easily find that the observations we documented in Section 3 are still observed. For instance, there are two breaks for employment of the high-skill occupation group when the occ1990dd classification is used: between 1982 and 1983 and between 2002 and 2003. Similar breaks are also observable in Figure A.2b, total hours of the high-skill occupation group constructed using the occ1990dd classification. Hence, data constructed using the conversion factors still outperform other data sets.

\(^{34}\)When aggregating individual data, we use the earnings weight (earnwt) that should be used in analyses of employment and hours/weeks worked as well as earnings.
Figure A.2a: Employment, Low-Skill Occupation

Figure A.2b: Total Hours, High-Skill Occupation

Figure A.2: The March CPS

Figure A.3a: Total Hours, Middle-Skill Occupation

Figure A.3b: Total Hours, Low-Skill Occupation

Figure A.3: The March CPS

Table A.1: Standard Deviations of the Detrended Employment: March CPS

<table>
<thead>
<tr>
<th>Occupation Group</th>
<th>High-Skill</th>
<th>Middle-Skill</th>
<th>Low-Skill</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) 71-83</td>
<td>(2) 84-10</td>
<td>(1) 71-83</td>
</tr>
<tr>
<td>Conversion Factors</td>
<td>1.01</td>
<td>.93 (0.92)</td>
<td>2.34</td>
</tr>
<tr>
<td>occ1990dd System</td>
<td>2.49</td>
<td>1.34 (0.55)</td>
<td>2.31</td>
</tr>
<tr>
<td>Raw Data</td>
<td>2.41</td>
<td>1.01 (0.42)</td>
<td>2.25</td>
</tr>
</tbody>
</table>

Note: 1. All numbers are multiplied by 100.
2. Numbers in parentheses are the ratios between (2) and (1).