Deutsches Institut für
Wirtschaftsforschung

DIW BERLIN

www.diw.de

SOEP The German Socio-Economic Panel Study

SOEPpapers
on Multidisciplinary Panel Data Research

318

Aysen Isaoglu

Occupational Affiliation Data and Measurement Errors in the German Socio-Economic Panel

Berlin, September 2010

Occupational Affiliation Data and Measurement Errors in the German Socio-Economic Panel

Ayşen İsaoğlu*

European University Institute
&
Luxembourg Income Study

This paper shows that there are severe measurement errors regarding the occupational affiliations in the German Socio-Economic Panel. These errors are traced back to the survey structure: in years where occupational information is gathered from the entire employed population instead of only from those declaring job or labor market status changes, average occupational mobility is around five times higher. In order to construct reliable occupational affiliation data, a correction method based on related job or labor market status changes is proposed. The corrected occupational mobility patterns are then analyzed for different samples.

JEL classification. C41, C81, J62.

Keywords. Measurement Errors, Occupational Mobility, Panel Data.

*I would like to thank Silke Anger, Ana Cardoso, Hartmut Lehmann, Rick van der Ploeg, Frank Vella, Sanne Zwart and participants of the 9th IZA European Summer School in Labor Economics for useful discussions and constructive comments.
E-mail address: aysen.isaoglu@eui.eu

1. Introduction

Occupational affiliation data is important for two growing aspects of labor economic research. The first is the determination of wage growth. The second is the analysis of worker turnover. However, reliability of data on occupational affiliation is known to be an issue (e.g. Mellow and Sider (1983), Murphy and Topel (1987), Mathiowetz (1992), Polivka and Rothgeb (1993), Neal (1999), Kambourov and Manovskii (2004*a*), Moscarini and Thomsson (2008)). This paper shows that measurement errors concerning occupational affiliations are severe in the German Socio-Economic Panel (SOEP). Throughout, the focus will be on occupational mobility at the individual level as it allows for displaying data inconsistencies in the clearest way. After discussing the sources of the measurement errors, a correction method based on job or labor market status changes is presented. Finally, the corrected average occupational mobility measures for different samples and occupational classifications are discussed.

Reliable data on occupational changes is crucial to analyze the contributing factors of wage growth. While many studies (e.g. Neal (1995), Parent (2000) and Dustmann and Meghir (2005)) argue that human capital is specific to the industry of employment i.e. that industry tenure has significant explanatory power on wage growth, Kambourov and Manovskii (2009), using Panel Study of Income Dynamics (PSID), provide evidence of considerable returns to occupational tenure. In fact when an individual's occupational experience is taken into account, his/her tenure in an industry or with an employer is found to have little importance in explaining his/her wage. Hence, they conclude that human capital is occupation rather than industry or employer specific.

Occupational changes are also of interest in studies of worker turnover. After analyzing worker reallocation across employment states (e.g. Abowd and Zellner (1985), Blanchard and Diamond (1990)), across employers (e.g. Farber (1994), Fallick and Fleischman (2001)), across industries (e.g. Jovanovic and Moffitt (1990), Bils and McLaughlin (1992)), recent studies have focused on worker reallocation across occupations (see Moscarini and Vella (2003), Kambourov and Manovskii (2004*b*), Burda and Bachmann (2008), Moscarini and Vella (2008)). These studies argue that occupations at a detailed level provide the best information to the labor economist about the career changes. To see the importance of changes across detailed occupations, consider, for instance, the broad title *Professionals [2]*, where the number in square brackets denotes the respective code of the International Standard Classification of Occupations (ISCO-88). This entry includes both *Meteorologists [2112]* and *Chemists [2113]*. Clearly, one would not be able to identify the important career change of becoming a chemist after having worked as a meteorologist if the classification is not considered at a disaggregated level. Even at the three-digit level both of these occupations are named under *Physicists, Chemists and Related Professionals [211]*.

Data on occupational affiliation is known to be subject to measurement errors. This is not surprising as occupational classifications may contain hundreds to thousands of units. Cross-sectional errors in coding that are overlooked may become apparent only when longitudinal dimension is considered. Therefore, one of the most obvious ways to investigate the reliability of occupational affiliations is to analyze occupational mobility patterns.

Plots of worker turnover across occupations using the data provided by SOEP exhibit a suspicious pattern over the last two decades. The fraction of workers changing occupation at annual frequency alternates recurrently between around 7 and 45 percent. These percentages are for the four-digit ISCO-88, which is constituted of 390 distinct occupational units. Even at the one-digit level, which only has 9 different occupational groups, the percentages are around 5 and 25 respectively. In this study it is shown that this pattern is mainly driven by the survey structure: years with high average occupational mobility coincide with the years in which the occupational information is gathered from all workers. In the years with low values, the occupational information is gathered only from respondents who declare that they have experienced a job or labor market status change.

To obtain more accurate occupational affiliation data, a correction method based on other reported job or labor market status changes is used. The rationale is that an occupational change is likely to be accompanied by a change of employer, position in the company, industry etc. Similar filters are also used by e.g. Moscarini and Thomsson (2008). This method clearly corrects the unacceptably high average occupational mobility found in years where every worker was interviewed about their occupation. The alternating pattern in the average occupational mobility disappears after correction which validates the claim that a substantial part of the measurement error stems from the structure of the survey.

Results are presented for two measures of average occupational mobility that are commonly used in the literature. The first measure considers a worker as a "mover" if he/she declares a different valid occupational code in two consecutive periods in which he/she is employed (see Moscarini and Thomsson (2008), Burda and Bachmann (2008), Moscarini and Vella (2008)). The second measure also considers switches after non-employment spells, i.e. if an individual is employed in the current period, but was not employed in the previous period, a switch in his/her occupation will be recorded if he/she reports a current occupation different from the one he/she reported when he/she was most recently employed (see Kambourov and Manovskii (2004*b*)).

Average occupational mobility at annual frequency is ranging from around 4.5 percent to 7 percent over the last two decades, depending on the sample and the classification choice. There is no trend, but strong procyclicality is found to be robust across different samples. Only when changes after non-employment spells are also considered females are more mobile on average than males. This is expected since females have more intermittent careers and after non-employment periods workers in general are more likely to change occupations. Interestingly, workers with at least a college degree are found to be more mobile on average in comparison with other educational groups. Not surprisingly, workers younger than 40 have a higher occupational mobility on average which is also driving the overall procyclicality. The inclusion of workers from the former German Democratic Republic raises the mobility levels significantly, especially when changes after non-employment spells are also considered. Adding government sector or self-employed workers to the sample does not have significant impact on the observed occupational mobility patterns. The average occupational mobility levels are increased slightly when part-time workers are included. This is found to be mainly driven by females joining the employment pool after non-employment.

There are other studies presenting findings on occupational mobility in Germany that, like this study, use individual level data and disaggregated occupational classifications. Very recently, Burda and Bachmann (2008) analyze the extent and the dynamics of structural change in western Germany using the Institute for Employment Research (Institut für Arbeitsmarkt und Berufsforschung (IAB)) dataset. They compute worker flows across occupations and sectors over the period 1975-2001. Occupational mobility levels and cycles presented in their study are very similar to this study's findings. Kambourov and Manovskii (2004*a*) also use the IAB dataset but for the period 1975-1995. For a sample with the same characteristics and for the common time span 1985-1995, albeit with a different occupational classification, it is found that the patterns of occupational mobility are very similar to the ones in this study. However, there is a difference in levels (11 percent versus 7 percent). In an other study, Zimmermann (1999) analyzes the period 1984-1991 using the SOEP. An exhaustive set of tables on average job and occupational mobility is provided, however measurement errors are not discussed. A direct comparison of the results presented in Zimmermann (1999) and the current paper is unfortunately not possible since the codes used in Zimmermann (1999) are no longer available in the SOEP. However, the average occupational mobility is about twice as high of this study's findings when corrected affiliation data is used for the same sample.

As mentioned before, measurement errors in occupational affiliation data are also an issue for other datasets. For instance, Kambourov and Manovskii (2009) document in detail the measurement errors regarding the occupational and industry affiliations in the PSID for the period 1968-1993. They compare the original occupation and industry affiliation data, i.e. coded at the time of the survey, to retrospectively coded data. The latter data files, namely, the Retrospective Occupation-Industry Supplemental Data Files became available in 1999 and include a retrospective assignment of three-digit 1970 Census codes to the reported occupations and industries for the period 1968-1980. There is stark disagreement between the originally and retrospectively assigned codes for the same individuals. For the period 1976-1980 two-digit occupational mobility levels in the retrospective files are found to be twice as small than the ones obtained in the original files.

Similarly, for the annual March files of the Current Population Survey (CPS) dataset Murphy and Topel (1987) and Kambourov and Manovskii (2004*a*) show strong evidence for classification errors in occupations. In the CPS, each household is interviewed once a month for four consecutive months, then removed from the sample for eight months and again interviewed for another four months. Thus, any household is present in the survey for eight (non-consecutive) months in total. Occupational and industry information is gathered monthly, regarding workers' labor force activity in the week prior to the survey. Additionally, in March of each year workers present in the CPS sample are given a supplemental questionnaire in which they are asked to describe their longest job last year. Kambourov and Manovskii (2004*a*) provide convincing evidence that annual data from the March files does not measure annual mobility correctly. Due to the rotation of the panel, this data merely measures mobility over a couple of months' period. Recently, Moscarini and Thomsson (2008) employ the CPS data at the monthly level in order to avoid this problem. They exploit the monthly longitudinal structure to derive more accurate occupational mobility data.

Although the focus of this study is the occupational affiliations, it should be pointed out that the industry affiliations in the SOEP (two-digit Nomenclature des Statistiques des Activités Economiques de la Communauté Européenne (NACE) and two other codes in the Cross-National Equivalent Files) are also measured with error as the information on the industry the worker is in was gathered through the same procedure as for occupational information. Moreover, any measure derived from occupational or industry affiliations is also contaminated with measurement errors such as International Socioeconomic Index of Occupational Status (ISEI), Magnitude Prestige Scale (MPS), Treiman Standard International Occupational Prestige (SIOPS) and Erikson Goldthorpe Class Category (EGP).

The next section describes the characteristics of the SOEP and the Section 3 provides information on the occupational classifications. Section 4 discusses the measurement errors and Section 5 explains the proposed correction method. Section 6 presents and discusses the corrected occupational mobility measures. Section 7 concludes. The Appendix provides a detailed description of the data correction and the related properties of the sample.

2. German Socio-Economic Panel

The SOEP is a nationally representative longitudinal survey of persons and private households which started in the Federal Republic of Germany (FRG) in 1984 with around 12,000 respondents (Wagner, Frick and Schupp (2007)). The target population represented in the SOEP was the entire residential population of the

FRG. Initially there were two samples, namely *Residents in the FRG* and *Foreigners in the FRG*. The first sample covers persons in private households with household heads who do not belong to the main foreigners groups of guestworkers, whereas the second considers the private households where the household head is from Greece, Italy, Spain, Turkey and former Yugoslavia. The SOEP expanded to the former German Democratic Republic (GDR) in June 1990 and since then the residential population in the former GDR is also represented (Haisken-DeNew and Frick (2003)).

The SOEP has various advantages and disadvantages for studying labor market transitions. The primary advantage is, next to transitions across the labor market status i.e. employment, unemployment or being out of labor force; transitions across firms, within firms, industries and occupations are also collected. Moreover information on the exact timing of these transitions is gathered either via explicitly asking for the month and year of the change or via questions based on a calendar.

A second advantage of the SOEP is the consistency of the survey questions. The central aim of this panel study is to collect representative micro-data on persons and households in order to measure stability and change in living conditions. Hence, changes in the questionnaires are minimized.

An additional advantage of the SOEP is that generated variables are also provided next to the direct responses from the surveys for some variables. These generated variables are more reliable since they are constructed using several cross-checks. As suggested by Haisken-DeNew and Frick (2003), generated variables are used instead of the direct survey responses in this study when both are available.

There are also disadvantages to using the SOEP. Compared to other datasets, such as the IAB, representing a two percent sample from the German social security records, there are relatively few observations. Moreover, as the SOEP is a survey, information is collected on a voluntary basis which makes it prone to suffer from attrition. The representativeness of the SOEP sample is addressed in several ways. All household members are interviewed individually once they reach the age of 16. Hence, the next generation is automatically included. In case of residential mobility, the person is followed within the country. Although this might lead to over- or underrepresented geographical areas, it does not affect other properties of the sample such as gender, age and family distribution.[1] Third persons moving into an existing SOEP household are surveyed even in case of subsequently leaving that household. Persons and households which could not be successfully interviewed in a given year are followed until there are two consecutive temporary drop-outs of all household members or a final refusal. In the case of a successful interview after a drop-out, there is also a small questionnaire including questions on central information which is missing for the drop-out year. Addresses are kept up to date by the field work agency throughout the entire year in order to be informed about residential mobility.

The analysis in this study is based on 21 waves that cover the period 1984-2004. The base sample consists of full-time employed males and females, aged between 18-65, members of the *Residents in the FRG* and *Foreigners in the FRG* samples, not receiving education or training, not dually employed, not self-employed or belonging to a household with a self-employed member, not working in the government sector. Observations for individuals who reported to be living in the former GDR in 1989 and who moved to the former GDR after the unification are also excluded.[2] Additional sample specifications will be discussed and analyzed in Section 6.

3. Occupational Classifications in the SOEP

The SOEP provides several occupational classifications. This study focuses on the "Klassifizierung der Berufe (KldB)", which is the national coding system of the German Federal Statistical Office, and the International Standard Classification of Occupations (ISCO-88). The KldB is provided at the four-digit level. The 2,287 occupational unit groups can be aggregated to units of 369, 88, 33 and 6. The ISCO-88 is a nested classification of occupations at the four-digit level. The one-digit distinguishes 9 major groups, which have 28 major subgroups, 116 minor groups and 390 unit groups. Classification at the four-digit level thus corresponds to 390 different occupations (ILO (1990)). The four-digit KldB and ISCO-88 classifications provide highly detailed occupational information. A third classification in the SOEP is in the Cross-National Equivalent File (Burkhauser, Butrica, Daly and Lillard (2000)), referred to as CNEF code in this study. Although less detailed (101 occupational

[1]Note that the panel structure together with the follow-up of individuals instead of addresses (where the latter is the case in the CPS used by Moscarini and Thomsson (2008)) allow taking into account occupational changes that are accompanied by geographical changes as well.

[2]See the Appendix for a detailed description of the employed sample.

units), this file consists of equivalently defined variables to allow for comparison of the PSID, the SOEP, the British Household Panel Study (BHPS) and the Canadian Survey of Labour and Income Dynamics (SLID). As this code is derived from the other occupational classifications, it is not discussed separately here.

The KldB and the ISCO-88 are present in the SOEP for all periods under the investigation. However, occupational information is not asked each year to the whole survey population. Instead, in 1985, 1986, 1987, 1988, 1990, 1992, 1994, 1996, 1999, 2001 and 2003 only respondents who declared a job or labor market status change was surveyed. In the rest of the years, the whole population was surveyed via a direct question:

> What is your current position/occupation? *Please give the exact title. For example, do not write "clerk", but "shipping clerk"; not "blue-collar worker", but "machine metalworker". If you are engaged in public employment, please give your official title, for example, "police chief" or "lecturer". If you are an apprentice or in vocational training, please state the profession associated with your training.*

In the years that the question is asked only to people who experienced a job or labor market status change, the previously declared occupation is coded in the absence of a job or labor market status change.

A recoding of the occupational affiliations based on the original survey responses took place in 2002 (see Hartmann and Schuetz (2002)). The main reason for the update was to replace the outdated ISCO-68 with the ISCO-88. Based on the original survey answers, occupational affiliations were recoded retrospectively according to various criteria. First, recoding was done using the national coding system KldB. These codes were then translated into ISCO-88 by an algorithm. If the respondent provided information referring to distinct occupations in his/her answers, the first mentioned occupation was taken unless information regarding the second occupation was more precise. When the respondent did not provide sufficiently specific information to distill an occupational affiliation, also information such as industry branch, training and the job position was taken into account to decide on what his/her occupation is. If this was still not informative enough to determine the occupational category of the respondent, then the following two rules applied according to the source of ambiguity. If the information on the content of the occupation was not

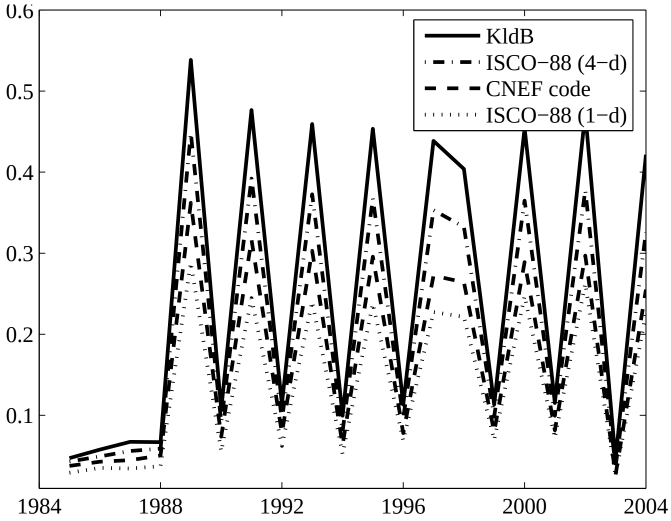

FIGURE 1. Occupational mobility based on the original data considering the national code (KldB), four-digit ISCO-88 (ISCO-88 (4-d)), CNEF code and one-digit ISCO-88 (ISCO-88 (1-d)).

sufficiently specific to fit a single category, the category more frequently observed in the data was chosen. If the information was only sufficiently specific to determine the category of the occupation, the occupation in this category with the lowest qualification level was chosen. For 96.4 percent of the respondents the information was sufficiently specific to unambiguously generate occupational codes (87.2 percent without any additional information and 9.2 percent with additional information). Only for the remaining 3.6 percent of the cases, the last two rules had to be taken into consideration.

4. Measurement Errors

There are severe and unambiguous measurement errors in the occupational affiliations in the SOEP. Figure 1 depicts the average occupational mobility over the last two decades for the base sample. Although in the figure only occupational transitions from employment-to-employment are considered, the picture is similar when changes after non-employment spells are also taken into account. Since the occupational changes are of interest, the first wave is lost.

Figure 1 is self-alerting as an evidence of measurement errors in the data. For all classifications available in the SOEP, average occupational mobility changes from

5-7 percent one year to 25-55 percent for the next year and then back again to 5-7 percent and this in a repeated manner.

The measurement errors in occupational mobility arise regardless of the disaggregation level. As mentioned above, the KldB considers 2,287 different units where as the one-digit ISCO-88 considers only 9 units.[3] Although measuring occupational mobility with the one-digit ISCO-88 lowers the peaks from 35-45 percent to 20-25 percent, the dented pattern remains.

The reason for the observed average occupational mobility patterns is clear: most of the errors are generated by the structure of the survey. One can clearly see that the years 1989, 1991, 1993, 1995, 1997, 1998, 2000, 2002 and 2004 with a high occupational mobility are also the years in which all respondents, independent of whether they have experienced a job or labor market status change, are asked to declare their occupation in detail. Apparently, asking this question without any dependence on other changes is vastly generating spurious changes.

One could argue that the peaks reflect accumulated occupational changes over subsequent years. When individuals do change occupation but fail to report a job or labor market status change, the occupational change is counted in the following year in which all individuals are surveyed. This would imply that when occupational information is asked from all respondents in two subsequent years, average occupational mobility in the second year should be about half of the value found in the first year. In 1997 and 1998 all respondents are asked about their occupations regardless of any job or labor market status change, nevertheless, magnitudes of the average occupational mobility are found to be similar. This suggests that it is mostly the survey design instead of the accumulated occupational changes that drives the pattern.

There are other well-known sources of measurement errors in occupational affiliations. It could be the case that respondents are explaining their tasks in an unclear way or that the coder generates the error while coding. Mathiowetz (1992) presents an experiment in which coders are asked to assign occupations based on company records and respondent records independently for the same sample. The disagreement rate is found to be 48 percent at the three-digit level. Clearly, in the

[3]When armed forces are also included, it becomes 10 units as armed forces are classified separately in the SOEP.

SOEP the scope of this kind of errors is larger in years when occupational information is collected from all respondents. Errors created at the coding stage can be minimized by retrospective checks on the same individual (see Kambourov and Manovskii (2004*a*)). However, the 2002 recoding of occupational codes in the SOEP mentioned above did not take advantage of this. The 97 percent precision of the recoding thus only relates to cross-sectional inferences. At the longitudinal dimension, errors are not specifically addressed. This explains the fact that there are many instances in the data where respondents are coded in two different (quantitatively) but very similar (qualitatively) occupations. For instance, consider the ISCO-88 codes. Someone who had declared an occupational change in 1996 and hence had been asked for the new occupational information had been coded as *Secretary [4115]*. The following year, when the entire population was asked the question related to their occupations, although she did not declare any kind of job or labor market status change over the last year, she was coded as *Stenographer or Typist [4111]* and the year after, where again the whole population was surveyed, without experiencing any job or labor market status change, she was coded *Philologist, Translator, Interpreter [2444]*. This result is probably driven by a combination of inconsistent response behavior and the coding error. This special case is also a good example to explain the differences in levels among the different classifications in Figure 1. In the example, the first "highly likely spurious" change from *[4115]* to *[4111]* would not be observed if one was considering the occupational changes at the three-digit level, the latter change is observed at all levels. Clearly, the more detailed the occupational code, the more prominent the measurement errors are, although some of these coding errors occur at all levels.

Evidence from the PSID recoding suggests that the measurement errors would indeed have been less severe if the recoding of 2002 would have used retrospective checks on the same individual. The PSID used one-digit occupational codes in 1968-1975, two-digit occupational codes in 1976-1980 and three-digit codes after 1981. In 1986, the PSID analysts started working on the 1968-1980 files in order to maintain three-digit occupational codes for the whole period of the survey, including the years before 1981. To create three-digit codes, original material, which was also used to create the one- and two-digit codes in the past, was used. In 1999 the Retrospective Occupation-Industry Supplemental Data Files was released. Kambourov and

Manovskii (2004*b*) find a considerable disagreement between the originally coded and the retrospectively corrected files. Occupational mobility in the latter for the period 1981-1994 is more than twice as small than the mobility obtained from the originally coded occupations. In the retrospectively corrected PSID files, all occupational information for each respondent across all required years was coded by the same analyst before moving to another respondent. In this way, the analyst also used the past and future information on the occupation of the respondent which obviously leads to a more consistent occupational history. In the next section, a similar approach is followed to correct occupational affiliation data in the SOEP.

5. Identifying the Genuine Occupational Changes

To reduce the measurement errors in occupational affiliations in the SOEP, a correction method that uses job or labor market status changes is followed. This is the most straightforward way of correcting the data since for almost half of the waves occupational information is asked only when respondents declare a job or labor market status change. Using such changes as a condition when correcting the data imposes the structure that is lost in years when the occupational coding is asked to all respondents. Moreover, it is unlikely to observe a genuine occupational switch without any other labor market situation change for a worker. Kambourov and Manovskii (2004*b*) show that in the PSID, 80 percent of the one- and two-digit occupational switches in the Retrospective Files are accompanied by either an employer or a position switch. The idea of the correction method is therefore to consider occupational changes genuine if they are accompanied by other job or labor market status changes.

Another justification of pursuing this approach follows from Polivka and Rothgeb (1993). Asking whether job or labor market status changes occurred starting from the beginning of the previous year is similar to using dependent coding. In the former, respondents are asked whether there were any changes, while in the latter respondents are confronted with their occupation in the previous period and are asked whether this is still their occupation. Polivka and Rothgeb (1993) analyze a proposed change in the survey structure of the CPS. When all respondents were asked to report their occupation, the average occupational mobility was 39 percent, whereas when dependent coding was used, it dropped to about 7 percent. In addition, an external consulting firm gave 5.9 and 7.4 percent as bounds for average occupational mobility. The conclusion was that using dependent coding leads to more accurate estimates. The similarities of dependent coding and asking about job or labor market status changes suggest the use of the latter to identify genuine occupational changes.

In each wave of the SOEP, the respondents are asked to state the changes in the "job situation" since the beginning of the previous year. If there is any change, they are asked to give information on the type and timing of the change such as whether the respondent has entered employment for the first time, started paid employment again after not being employed for a while, started a new position with a different employer, became self-employed or changed positions within the same company. Any occupational change accompanied by one of the job or labor market status changes above is considered as genuine. Without such a change the previous occupational code is kept. How this method is implemented in practice is discussed in the remainder of this section.[4]

To motivate the correction method, questions regarding job situation changes as well as questions regarding occupation and industry information are presented in Table 1. They are taken from the 2001 and 2002 surveys. Note that the latter year is a "peak year" and the former not.

From question 23, one can see that if the respondent declares in 2001 that he/she had not experienced a job situation change, then he/she is not asked for occupation or industry information. However, in 2002, regardless of the job situation change his/her occupational and industry information is asked.

The method to identify genuine occupational changes in the related calendar years consists of three steps. Before going into details, these steps can be summarized as follows. First, it is checked whether the respondents have changed their job or started a new job after the beginning of the previous calendar year (question 23 in Table 1). In case of no reported change, the previously coded occupation is kept. Second, if a change in job or labor market status took place, then the type and the exact timing of the change is retrieved (questions 24 and 25). Third, the

[4]Detailed information on the changes made to the data and imputation methodology is presented in the Appendix.

23. Did you change your job or start a new one after December 31, 1999? (2000 *in the 2002 questionnaire*)

- yes ☐

- no ☐, skip to question 37 (skip to question 30 *in the 2002 questionnaire*)

24. When did you start your current position?

2000, in the month ☐☐ (2001, in the month ☐☐ *in the 2002 questionnaire*)

2001, in the month ☐☐ (2002, in the month ☐☐ *in the 2002 questionnaire*)

25. What type of an employment change was that?

In the case that you have changed positions several times, please pick the appropriate reason for the most recent change.

- I have entered employment for the first time in my life ☐
- I have started up with paid employment again after not having been employed for a while ☐
- I have started a new position with a different employer ☐
- I have become self-employed ☐
- I have changed positions within the same company ☐

30. What is your current position/occupation?

Please give the exact title. For example, do not write "clerk", but "shipping clerk"; not "blue-collar worker", but "machine metalworker". If you are engaged in public employment, please give your official title, for example, "police chief" or "lecturer". If you are an apprentice or in vocational training, please state the profession associated with your training.

35. In which branch of business or industry is your company or institution active for the most part?

Please state the branch as exactly as possible, for example, not "industry", but "electronics industry"; not "trade", but "retail trade"; not "public service", but "hospital".

37. Since when have you been working for your current employer?

If you are self-employed, please indicate when you started your current work.

Since, month ☐☐ year ☐☐☐☐

Table 1. Questions used in identifying genuine occupational codes in the 2001 and 2002 surveys. Differences in the 2002 questionnaire are mentioned between parentheses. Note the different implications of a "no" to question 23.

occupational information regarding to that change in the data is kept unchanged but deployed to the relevant calendar year if necessary (question 30).

In the first step, the occurrence of job changes are analyzed. To increase reliability, the generated variable is used. This variable shows whether the respondent is not employed, employed without a job change or employed with a job change. When the respondent is not employed the occupation is left missing, when the respondent is employed without a job change, the last reported occupation is kept. In case of a change in job status it is necessary to further analyze the occupational information.

The next step deals with identifying the calendar year of the change. This is important for two reasons. First, in contrast to other micro datasets where all interviews are held during a particular week or month, the SOEP survey is conducted all over the year. To have a consistent overall picture, it is important that for all respondents the same 12 month period should be used and the calendar year is the obvious candidate. Second, deploying changes to exact calendar years makes it possible to relate worker reallocation with macroeconomic variables from other sources. Almost 90 percent of the survey is held in the first four months of the calendar year. Therefore, a large fraction of job situation changes reported in a given year correspond to the previous year. After this recoding of the job situation changes according to the exact year of the change, as expected some individuals have multiple job situation and thus occupational changes in a given calendar year.

As a result of allocating changes to their calendar years, there are cases in which the respondent is not employed at the time of the survey in a given year and the year after declares a change considering the “previous year”. This raises the question whether to consider someone in the employment pool in a given calendar year when part of the year he/she is not employed. This choice obviously affects the occupational mobility. To reduce the scope for both under- and overestimation of occupational mobility, someone is considered “employed” if he/she works minimum 6 months in a given year. Respectively, relevant occupational codes and other variables, for instance, somebody becoming a government sector worker or self-employed with that job situation change, are also imputed. The results are not substantially altered when instead of 6 months, the minimum employed period is considered to be 3, 9 or 12 months.

Finally, after correcting the job situation change variables, new occupational codes are imputed. It is implicitly assumed that the occupational change took place when the job situation change took place. Double job situation changes is translated to only 41 occupational changes. When there is a double occupational change observed for the same calendar year, they are both counted for the aggregate occupational mobility measures.

A slight change in the survey questions in 1994 may have affected the occupational mobility measures. Before 1994, respondents were asked to declare *all* the job situation changes they have experienced from the beginning of the previous year until the current date of the survey. However since 1994 they are asked to declare *only* the last change. The data suggests that the change in the survey can be ignored while identifying the job situation changes. Out of 72,482 observations before 1994, there are only 119 observations for which multiple job situation changes are declared. Hence, ignoring multiple job changes when considering occupational mobility at annual frequency seems not to be problematic.

Since a substantial part of the current year information becomes only available in the following year, the last (incomplete) wave for every respondent is ignored unless he/she already reports a change in the first few months. For instance in 2004, for the last wave of the survey, an implausibly low level of occupational change is observed. This is mainly due to the fact that data for individuals who will declare a job situation, and possibly an occupational change for 2004 in 2005 are not available.

A similar correction method which is also considering occupational changes genuine depending on other provided information is also followed by Moscarini and Thomsson (2008) for the monthly CPS files. They employ four consecutive months to identify valid occupational changes between the second and third month. They thus do retrospective and retroactive checks on the same individual to minimize spurious changes. Sequences of four consecutive occupations that involve two transitions forth and back to the initial occupation and that do not correspond to changes in industry or class of workers or to active job search in the past month are considered suspicious. Using these filters such as active job search is attractive since the data is provided on a monthly basis. A high rate of transitions on a monthly level is more suspicious than on a yearly level. For annual data, it is more acceptable when an individual has four different occupations in four consecutive years. This is also valid for the active job search filter. For the SOEP data it is unfortunately not possible to employ industrial affiliation data as a filter. Note that information regarding the industry is asked in question 35 (see Table 1), i.e. in the same years as the question regarding occupations. The industry changes exhibit the same dented pattern as the occupational changes. Using them in identifying occupational changes will only introduce more noise.[5]

6. Occupational Mobility in Germany 1985-2003

There are two measures of interest for occupational mobility. The first measure considers an individual as a "mover" if he/she is employed in two consecutive years and reporting different occupations (hereafter, employment-to-employment). The second measure also considers occupational changes after a non-employment period. For instance, if an individual is employed in the current year, but was not employed in the previous year, a switch in his/her occupation will be recorded if he/she reports a current occupation different from the one he/she reported when he/she was most recently employed.

The corrected occupational mobility patterns are plotted in Figures 2 and 3 for the base sample discussed in Section 2, respectively for the two definitions of mobility mentioned above.[6]

From Figure 2 it can be seen that if one considers employment-to-employment changes only, occupational mobility averages to about 4.5-5 percent. As expected, the KldB which is more disaggregated leads to a higher average mobility than ISCO-88. Although there is no apparent trend, occupational mobility is clearly procyclical. Mobility was above average in 1989-1992, 1999 and 2001. The first period of high mobility is very likely to be related to the pre-unification economic boom and the unification itself. The trough in 1994 is expected to be the reflection of the 1993 recession.

If one also considers changes after non-employment spells, in general average occupational mobility rises to higher levels, see Figure 3. These higher levels reflect

[5]Figures for NACE, and one- and two-digits codes provided in the CNEF files are available from the author upon request.

[6]These figures might underestimate the true average occupational mobility as occupational changes at the worker level can only be identified when the worker also experience a job or labor market status change.

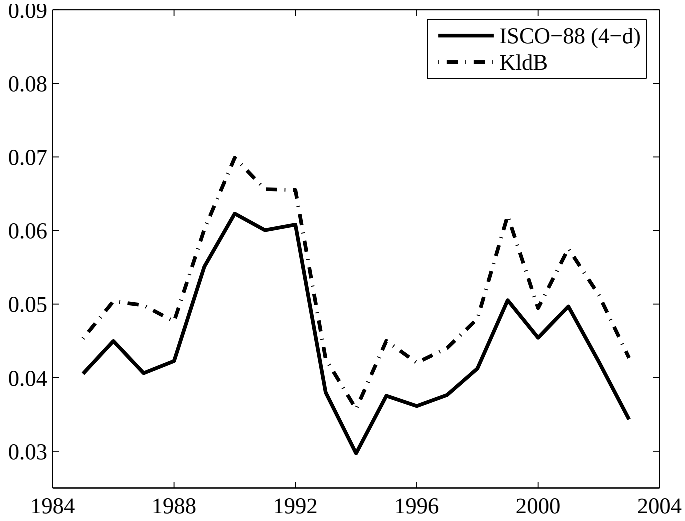

FIGURE 2. Occupational mobility considering employment-to-employment changes only.

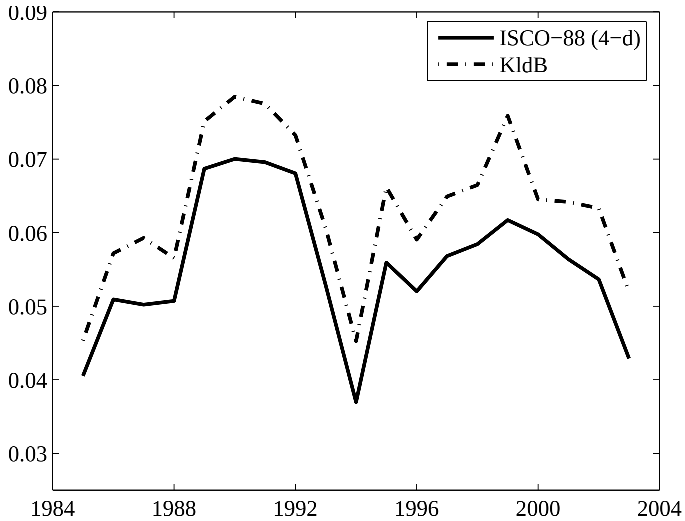

FIGURE 3. Occupational mobility including changes after non-employment spells.

the fact that after being non-employed, individuals are more likely to find work in an occupation different than their last. This can be due to, for example, loss of skills or a changing economy in which certain occupation appear and disappear over time. As before, the KldB classification leads to a higher average mobility. There is no clear trend and the cyclical pattern remains unchanged. One might argue that relatively higher levels after 1993 compared to Figure 2 reflect increasing higher unemployment rates in Germany.

Similar results are found by Burda and Bachman in their recent work (Burda and Bachmann (2008)). They use the IAB dataset to analyze worker flows across sectors and occupations in western Germany during the time period 1975-2001. The analysis considers 16 broad economic sectors and 128 different occupations. Occupational mobility considering employment to employment flows for males aged 30-49 during the period 1985-2000 ranges between 2 and 5 percent. They find peaks around 1990 and 2000 which coincides with Figure 2. They also find that average occupational mobility decreases with age and the probability of changing occupation is higher after a non-employment spell. However, for women instead of a higher they find a lower average occupational mobility.

The only other study which analyzes occupational mobility at a disaggregated level using SOEP is Zimmermann (1999). This study covers the period 1984-1991 for a sample of females and males, aged between 15 and 65. Individuals receiving vocational training and self-employed with their family members are dropped (see Page 311 and Table 12.3 of that study). The study presents general characteristics of the German labor market and partially deals with occupational mobility. An exhaustive set of tables containing information on job and occupational changes concerning different age groups, job status and educational levels is presented. The one- and three-digit ISCO-68 occupational codes are considered. Unfortunately, a direct comparison of the results presented in Zimmermann (1999) and the current study is not possible since the codes used in that study are not available anymore in the SOEP after the recoding that took place in 2002. Although a direct comparison is not feasible, it is still interesting to have a closer examination of the findings of Zimmermann (1999). In the study, measurement error issues are not addressed. As can be seen from Figure 1, there are only two years with "strange" spikes in the period 1984-1991. Since the average occupational mobility over time is not plotted and as only averages for the whole period are presented, the spurious changes in 1989 and 1991 might very well not be discovered. The reasonable occupational mobility levels of the first years further conceal what is going on in 1989 and 1991. Average occupational mobility is reported to be around 13 percent. When now four-digit ISCO-88 codes are considered for the same sample with the same characteristics, average occupational mobility for the uncorrected data is about 14 percent (17 percent for the KldB). With the corrected data the numbers are 5 and around 6 percent respectively. Since the three-digit ISCO-68 (1,506 occupational units) is more detailed than four-digit ISCO-88 (390 occupational units) and less detailed than KldB (2,287 occupational units), occupational mobility for that period is expected to be between 5 and 6 percent which is less than half of the reported value.

In Figure 4 the occupational mobility patterns are shown for different groups of gender, education and age using the base sample with four-digit ISCO-88 codes. The figures on the left hand side concern occupational changes with employment-to-employment changes only, the figures on the right hand side also allow changes after non-employment spells. Average occupational mobility for females and males are found to be similar when only employer-to-employer changes are included. When changes after non-employment spells are added, occupational mobility for females is higher especially after the unification. Occupational mobility for females also seems to be more volatile in general.

Three broad educational groups are distinguished in the figures, namely "high school and less", "high school and vocational" and "college and more". The first group considers individuals who have no school degree or only high school degree without any vocational training. The second group consists of individuals who successfully completed both high school and vocational training. Individuals in the last group have at least a college degree. It is surprising to see that individuals with a college degree are more mobile on average. One might have expected a lower occupational mobility due to occupation specific education that colleges provide.

The lowest row of figures shows the occupational mobility patterns for different age groups, more specifically below or above 40, the average age in the sample. As expected, older workers are less often changing occupations. The group with younger workers also shows clearer cyclical patterns. Apart from a pronounced drop around 1994, the occupational mobility of the older group seems to be unsensitive to macro-economic fluctuations.

To analyze the effect of the sample choice on occupational mobility patterns, different samples are used in Figures 5 and 6. For all figures, the base sample considered until now is extended with a particular group of workers, namely workers from the former GDR, government sector workers, self-employed and part-time workers. Again, occupational mobility is shown for both mentioned measures using four-digit ISCO-88 codes. First, the base sample is extended to include workers from

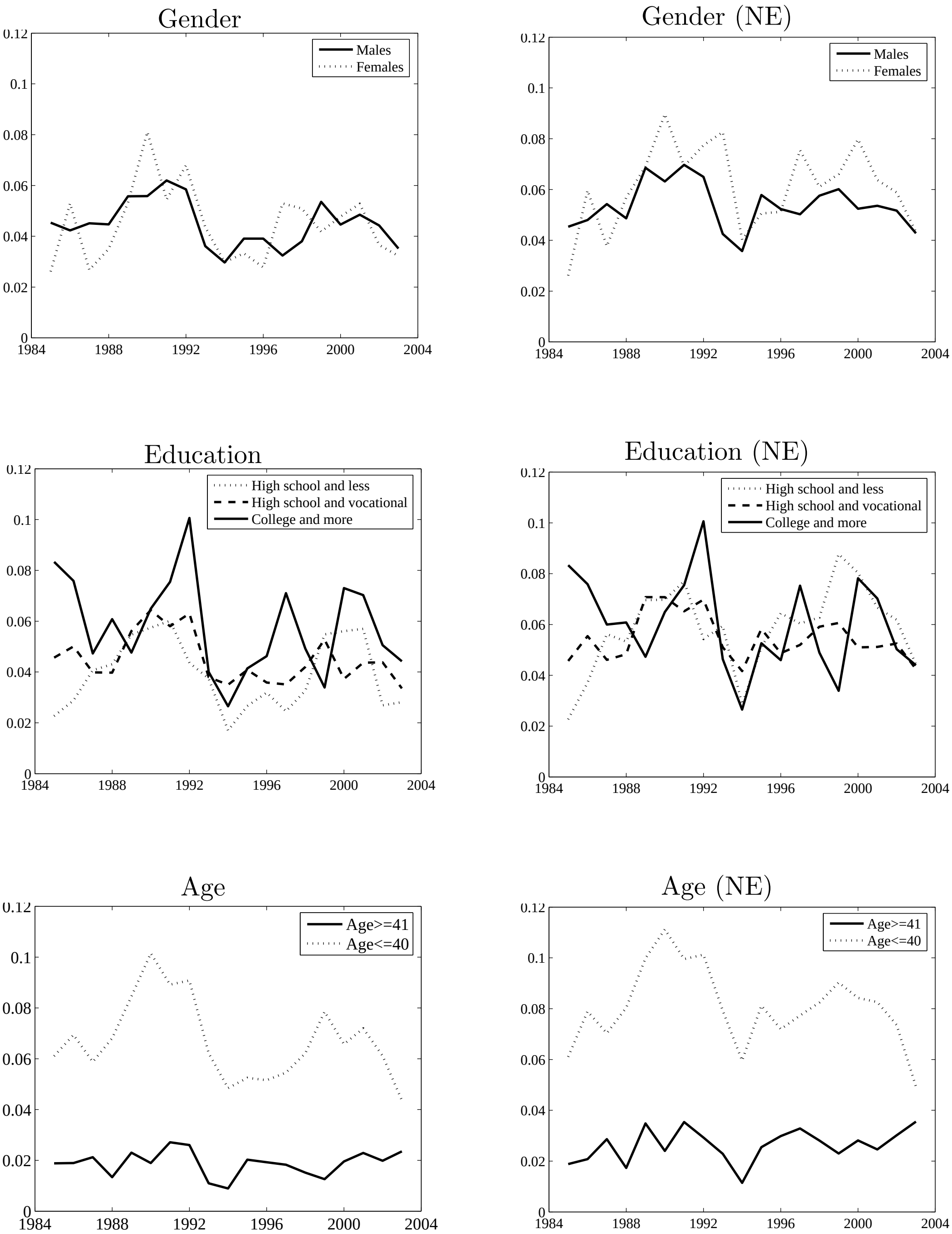

FIGURE 4. Occupational mobility for groups with different characteristics when considering employment-to-employment changes only and when including occupational changes after non-employment spells (NE).

the former GDR. For sake of consistency, the previously dropped individuals who

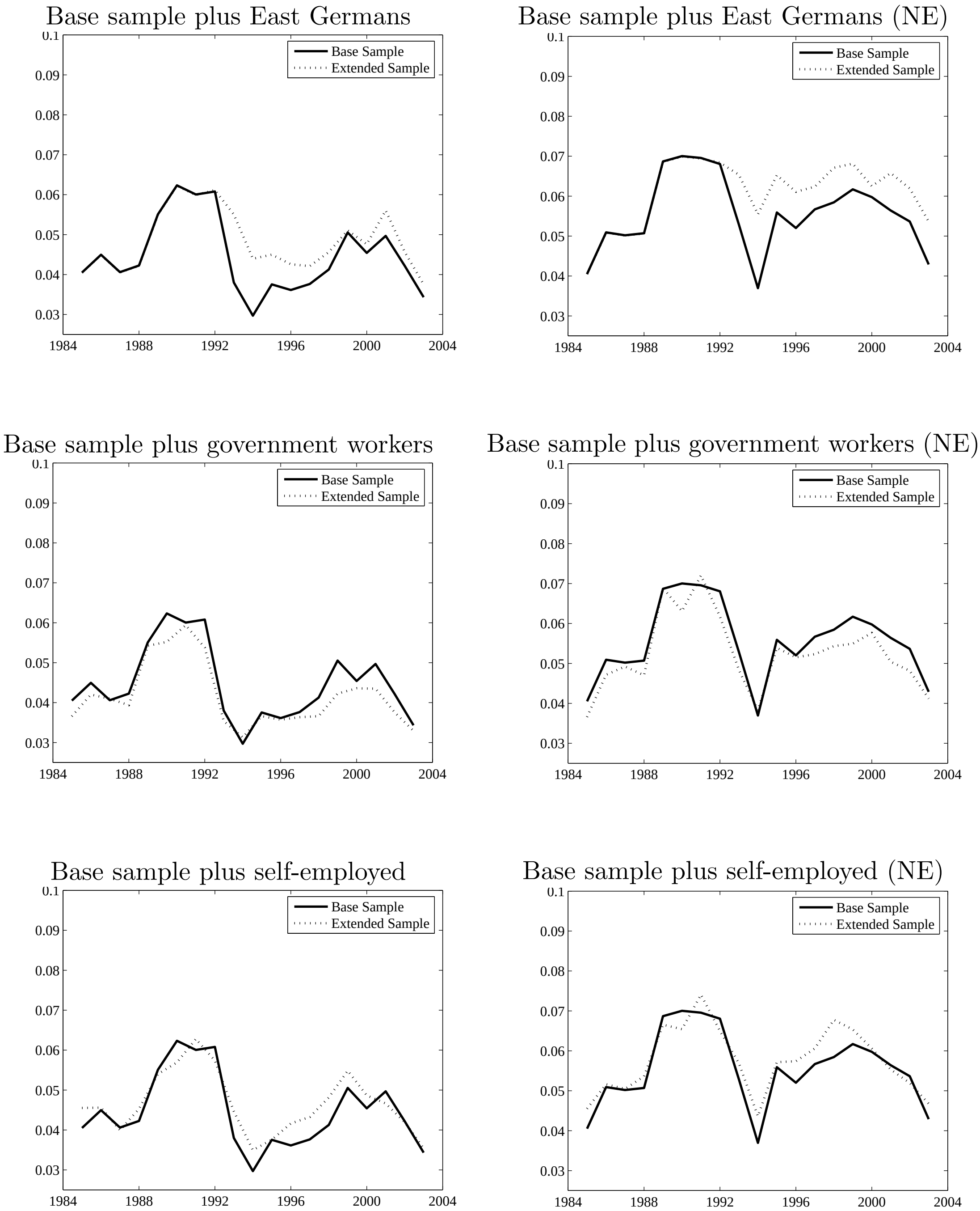

FIGURE 5. Base sample plus different groups when considering employment-to-employment changes only and when including occupational changes after non-employment spells (NE).

were living in the GDR prior to 1989 and the individuals who move there after unification are included. Although the SOEP started collecting data in the former GDR

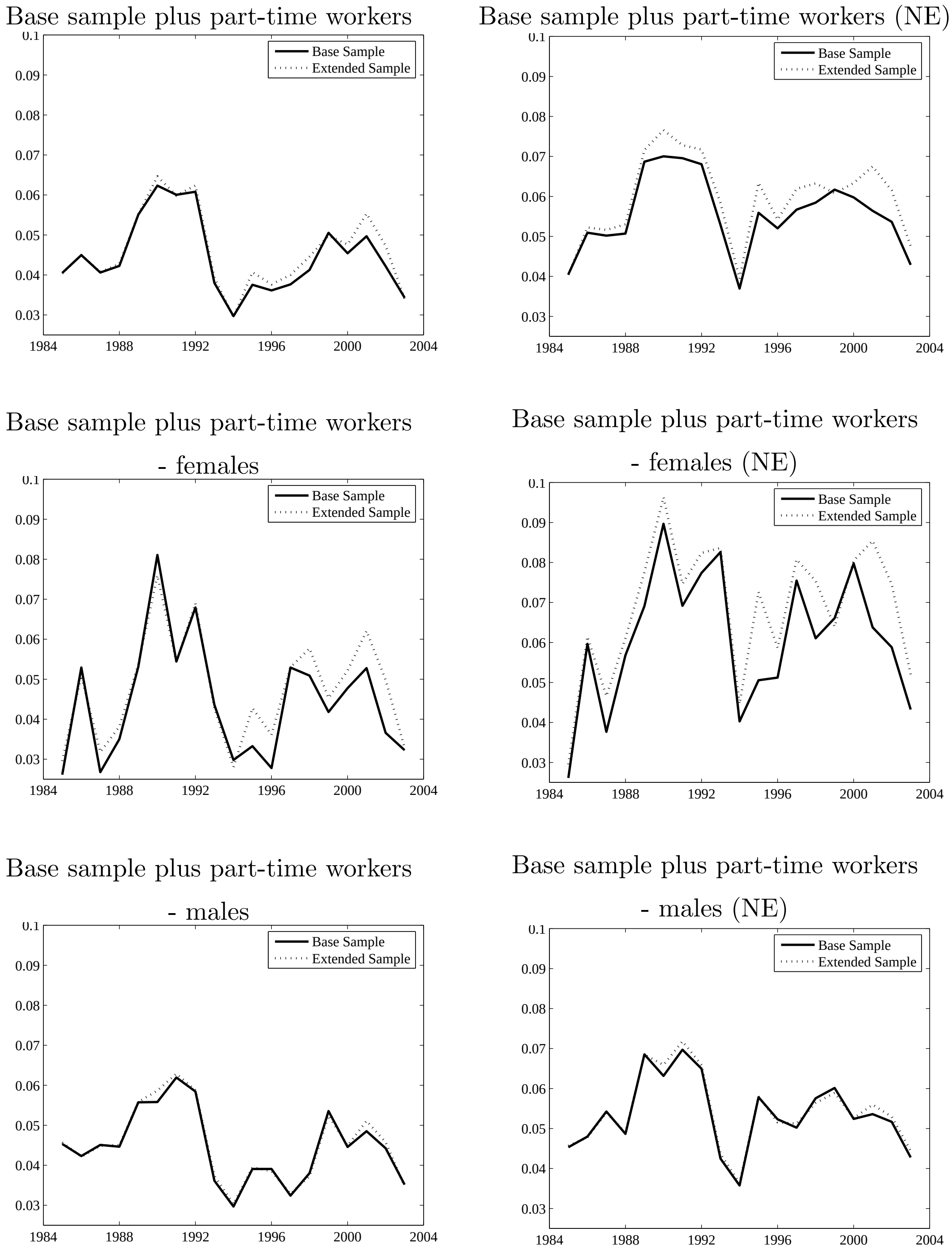

FIGURE 6. Base sample plus part-time workers when considering employment-to-employment changes only and when including occupational changes after non-employment spells (NE).

already in 1990, job and occupational information are only collected since 1992 so there is no data on occupational mobility for this sample before 1992. The difference in observed occupational mobility levels is stark after the inclusion of this sample. Especially when changes after non-employment spells are taken into account, consideration of workers from the former GDR translates to an almost constant level increase of one percent. The drastic decrease in 1994 is also mitigated considerably with the high occupational mobility levels of workers from the former GDR.

When one only focuses on workers from the former GDR, annual average occupational mobility drops from 9.5 percent in 1993 to around 4.5/5 percent in the period after 1997. This suggests that this group was in the process of occupational sorting right after the unification. Similar results are found for other transition economies. For instance, Sabirianova (2000) employing the Russian Longitudinal Monitoring Survey investigates the magnitude and the determinants of occupational mobility in Russia from 1985 to 1998. She finds that between 1991 and 1998, 42 percent of the employed changed their occupation, which is nearly twice the share of occupational movers in the previous six pre-transition years. In addition, she finds that the occupational flows were most intense during the first five years of reforms and that after 1996 the rate of occupational mobility began to fall. Her analysis concludes that the structural changes account for a substantial part of the increase in gross occupational flows. More recently, Campos and Dabusinskas (2003) analyze the process of occupational change in Estonia using data from the 1995 Estonian Labour Force Survey. They find that between 35 and 50 percent of wage earners changed occupations from 1989 to 1995. Again, they find that most of these occupational changes took place during the first years of the transition. They find that the typical change of occupations involved stepping down both the schooling and earnings ladders. They thus conclude that the process of occupational change was driven more by the transition itself than by individual workers choice.

The inclusion of government sector workers slightly decreases average occupational mobility, while the inclusion of self-employed leads to a slight increase. Although these sectors have quite different characteristics with respect to labor contracts, the occupational mobility patterns are not affected by the inclusion of either group.

Finally, in Figure 6 part-time workers are added to the base sample. Levels of occupational mobility are not affected much, except when changes after non-employment spells are also considered. Since a large proportion of part-time workers in Germany are known to be females, it is interesting to distinguish occupational mobility according to gender. It follows that the observed higher occupational mobility levels is due to the inclusion of part-time female workers. For male workers there is almost no effect.

7. Conclusion

This paper first presents unambiguous evidence for the existence of measurement errors in occupational affiliation data in the SOEP. These errors are caused by the structure of the survey. More specifically, gathering occupational information from all respondents independent of any other job or labor market status changes in certain years in addition to coding errors generates unacceptable spurious flows. Secondly, in order to minimize the measurement errors, the occupational data is corrected. The proposed method is based on considering occupational changes genuine only if they are accompanied with other job or labor market status changes. Thirdly, using the corrected codes, average occupational mobility patterns are presented for the last two decades. Depending on the disaggregation level of the used classification and the sample, occupational mobility averages to 4.5 to 7 percent. The pattern is found to be consistently procyclical for all samples and occupational classifications.

The particular survey structure does not only contaminate the occupational affiliation data. Industry affiliations and several social economic indices derived from occupation and industry information are also affected.

In a companion paper (see İsaoğlu (2010)), corrected occupational mobility patterns are analyzed in more detail. The panel structure of the SOEP and the corrected occupational mobility measures are exploited to identify the factors explaining the found patterns.

Bibliography

Abowd, J. M. and Zellner, A. (1985), 'Estimating Gross Labor-Force Flows', *Journal of Business and Economic Statistics* **3**(3), 254–283.

Bils, M. and McLaughlin, K. (1992), 'Inter-Industry Mobility and the Cyclical Upgrading of Labor', *NBER Working Paper Series* **4130**.

Blanchard, O. and Diamond, P. (1990), 'The Cyclical Behavior of the Gross Flows of the U.S. Workers', *Brookings Papers on Economic Activity* **1990**(2), 85–155.

Burda, M. C. and Bachmann, R. (2008), 'Sectoral Transformation, Turbulance, and Labor Market Dynamics in Germany', *IZA Discussion Paper* (3324).

Burkhauser, R., Butrica, B., Daly, M. and Lillard, D. (2000), 'The Cross-National Equivalent File: A product of cross-national research', *mimeo* .

Campos, N. F. and Dabusinskas, A. (2003), 'So Many Rocket Scientists, So Few Marketing Clerks: Occupational Mobility in Times of Rapid Technological Change', *William Davidson Working Paper* (552).

Dustmann, C. and Meghir, C. (2005), 'Wages, Experience and Seniority', *The Review of Economics Studies* **72**(1), 77–108.

Fallick, B. and Fleischman, C. (2001), 'The Importance of Employer to Employer Flows in the U.S. Labor Market', *Finance and Economics Discussion Series* (2001-18).

Farber, H. (1994), 'The Analysis of Interfirm Worker Mobility', *Journal of Labor Economics* **12**(4), 554–593.

Haisken-DeNew, J. and Frick, J. (2003), *DTC (Desktop Companion to the German Socio-Economic Panel Study)*, DIW, Berlin.

Hartmann, J. and Schuetz, G. (2002), Die Klassifizierung der Berufe und der Wirtschaftszweige im Sozio-oekonomischen Panel- Neuvercodung der Daten 1984-2001, Technical report, DIW.

ILO (1990), 'International Standard Classification of Occupations: ISCO-88', Geneva.

İsaoğlu, A. (2010), 'Worker Reallocation across Occupations in Western Germany', *SOEPpapers* (319).

Jovanovic, B. and Moffitt, R. (1990), 'An Estimate of a Sectorial Model of Labor Mobility', *Journal of Political Economy* **98**(4), 827–852.

Kambourov, G. and Manovskii, I. (2004*a*), 'A Cautionary Note on Using (March) CPS Data to Study Worker Mobility', *mimeo* .

Kambourov, G. and Manovskii, I. (2004*b*), 'Rising Occupational and Industry Mobility in the US: Evidence from the 1968-1993 PSID', *IZA Discussion Paper* (1110).

Kambourov, G. and Manovskii, I. (2009), 'Occupational Specificity of Human Capital', *International Economic Review* **50**(1), 63–115.

Mathiowetz, N. A. (1992), 'Errors in Reports of Occupation', *The Public Opinion Quarterly* **56**(3), 352–355.

Mellow, W. and Sider, H. (1983), 'Accuracy of Response in Labor Market Surveys: Evidence and Implications', *Journal of Labor Economics* **1**(4), 331–344.

Moscarini, G. and Thomsson, K. (2008), 'Occupational and Job Mobility in the US', *Scandinavian Journal of Economics* .

Moscarini, G. and Vella, F. (2003), 'Occupational Mobility and Employment Reallocation: Evidence from the NLSY79', *mimeo* .

Moscarini, G. and Vella, F. (2008), 'Occupational Mobility and the Business Cycle', *IZA Discussion Paper* (3369).

Murphy, K. and Topel, R. (1987), Unemployment Risk and Earnings: Testing for Equalizing Wage Differentials in the Labor Market, *in* K. Lang and J. Leonard, eds, 'Unemployment and the Structure of the Labor Market', New York, Basil Blackwell.

Neal, D. (1995), 'Industry Specific Human Capital: Evidence from Displaced Workers', *Journal of Labor Economics* **13**(4), 653–677.

Neal, D. (1999), 'The Complexity of Job Mobility among Young Men', *Journal of Labor Economics* **15**(2), 237–261.

Parent, D. (2000), 'Industry-Specific Capital and the Wage Profile: Evidence from the National Survey of Youth and the Panel Study of Income Dynamics', *Journal of Labor Economics* **18**(2), 306–323.

Polivka, A. E. and Rothgeb, J. M. (1993), 'Redesigning the CPS Questionnaire', *Monthly Labor Review* pp. 10–28.

Sabirianova, K. Z. (2000), 'The Great Human Capital Reallocation: A Study of Occupational Mobility in Transitional Russia', *Economics Education and Research Consortium Working Paper Series* (2K-11).

Wagner, G. G., Frick, J. R. and Schupp, J. (2007), 'The German Socio-Economic Panel Study (SOEP) - Scope, Evolution and Enhancements', *Schmollers Jahrbuch* **127**(1), 139–169.

Zimmermann, K. (1999), 'German Job Mobility and Wages', *IZA Reprint Series,* No. A-1/1999 .

Appendix. Details of the Data Cleaning Procedures

There are 134,182 observations of females and males aged between 17 and 65, not currently receiving education and/or vocational training, that belong to the "Residents in the FRG" or "Foreigners in the FRG" samples of the SOEP for the period 1984-2004.

Corrections regarding occupational affiliations in the data that are explained in detail in Section 5 are implemented after the "job situation change" variable is constructed.

Job situation changes are identified through the variable "erwtyp$$"($$ is the symbol used for the year in the SOEP, i.e. "erwtyp95" refers to the year 1995). It is a *generated* variable, so created after several cross checks, and it provides more consistent and reliable information than the direct survey responses. Unfortunately, this variable provides information only on whether there is a job situation change and not on the "type" and the "timing" of the change. The information on the "type" and "timing" of the changes from direct survey responses are combined with the information from the "erwtyp$$" variable. As can be seen from the survey questions in Section 5, "type" refers to information whether the respondent is in the labor market for the first time, comes back to employment after a non-employment period, changes job between employers, becomes self-employed or changes job within the same company. "Timing" refers to the exact month in a given year the change has realized. Accordingly, 769 observations are dropped as the "erwtyp$$" variable is missing.

As mentioned in Section 5, there was a change in the survey questions in 1994 which may affect the occupational mobility measures. Before 1994, respondents were asked to declare *all* the job situation changes they have experienced from the beginning of the previous year until the current date of the survey. However since 1994 they are asked to declare *only* the last change. The data suggests that the change in the survey can be ignored while identifying the job situation changes. Out of 72,482 observations before 1994, there are only 18 that show more than one job situation change for the same year. For these 18 observations, in order to be consistent with the data after 1994, only the last changes are considered. There are 119 observations for which two different changes are declared; one for the "current year" and one for the "previous year". In 38 of those cases the respondent did not participate in the survey or was not employed at the time of the previous year's survey, hence the occupational information regarding to that specific change is missing.

There are two obvious situations in which job situation change variables are suspicious. When there is a "change in the previous year" declaration and a "change in the current year" declaration made in the previous year that correspond to the same type of job situation change in the same month they are almost surely referring to the "same" change. In the whole sample there are 168 of such cases. There are also 82 cases where respondents declare two different job situation changes for the same month of the same year. Those are possibly but unlikely referring to two different changes. For instance, 63 of them refer to "come back to employment after a non-employment period" and "change job between firms" for the same month which suggest that respondents are simply providing extra information about their job situation, i.e. they have come back to the employment pool with a new employer. Therefore "change in the previous year" declarations are ignored.

After correcting the job situation change variable, new occupational changes are generated. This is done in two steps; the first step considers an observed change in occupational coding genuine only if it is accompanied with a job situation change. Here, also another survey question is used which inquires information on whether the respondent has *left* his/her job since the beginning of the previous year and if so when, in order to increase reliability in identifying occupational changes. The second step allocates the occupational changes to the exact year of the change. For the respondents who do not declare any change for the "current year" and for the "previous year" in the consecutive year, occupational information is kept as it is. Also for respondents who declare a change in the occupation for the "current year" the occupational change is kept. If the worker declares that he/she has experienced a change in the "previous year" and no change in the "current year", then this information is deployed to the previous year.

Obviously, there are also cases in which the respondent is not employed at the time of the survey in a given year and the year after declares a change considering the "previous year" (2,120 observations). This raises the question whether to consider someone in the employment pool in a given calendar year when part of the year he/she is not employed. Considering someone employed when he/she works only a small part of the year would lead to an underestimation of occupational mobility. Therefore someone is considered "employed" if he/she works minimum 6 months in a given year. 914 of the observations are then considered as employed in a given year although they have reported that they are unemployed at the time of the survey. For the rest of the 2,120 observations, the change is considered to be realized in the current year. Following the same criteria, 1,141 respondents that are recorded as working in a given year but in the consecutive year declared that they have left their job the previous year, are recoded as unemployed if the total time they are employed that year is less than 6 months.

Then, relevant occupational codes and other variables, for instance, somebody becoming a government sector worker or self-employed with that job situation change, are imputed.

After this recoding of the job situation changes according to the exact time of the change, there are 402 double job changes in a given year. In these cases, a job situation change is reported and the year after there is another change reported corresponding to the previous year. The "erwtyp$$" variable showed a change for both years although in fact both changes are realized in the same year. These double job situation changes is translated to only 41 occupational changes (61 for the sample that also considers changes after non-employment spells).

Consecutively, 276 observations regarding the individuals who moved to the former GDR and 281 observations for who used to live in the former GDR before unification are dropped. 21,186 observations of government workers and 17,023 observations for self-employed and their family members are dropped. Furthermore, 2,037 observations for dual employed and 6,880 for part-time workers are dropped.

After generating the binary variable that identifies the occupational changes, i.e. after using all the information SOEP provides, the first wave is dropped as the job situation change questions were not asked in 1984. However, information provided in 1984 on the occupation of respondents is used to find out the changes in 1985.

Moreover, the year that a respondent is not observed the consecutive year are not used unless he/she already declared an occupational change. The reason for that is again the fact that most of the job/occupational changes in a given year are declared in the consecutive year. That is also the reason why the last wave 2004 is dropped.

Finally, after deleting the 80 observations that are at age 17 (they were kept until this stage as they may provide information on occupations for some observations that are employed at age 18) the sample that is used for plotting occupational mobility consists of around 32,031 observations comprising employed individuals. There are small differences in the sample size depending on whether KldB or ISCO-88 is used. For the 19 years under consideration there is an average of 1,686 observations.