Schmollers Jahrbuch 125 (2005), 167 – 181
Duncker & Humblot, Berlin

The German Microcensus as a Tool for Longitudinal Data Analysis: An Evaluation Using SOEP Data*

By Edin Basic, Ivo Marek, and Ulrich Rendtel

Abstract

The German Microcensus is a rotating panel, where the units stay in the survey for four observations. Because of the very large sample size and the mandatory participation it appears to be a valuable data base for short duration analysis. However, the German Microcensus (MC) uses area sampling where participants are not followed if they leave the area. Consequently there is no information on participants after they move. We investigate how the use of the German Socio-Economic Panel (SOEP) can help to measure the non-coverage bias of the MC. Our methodology is evaluated for labour force flows. The results indicate that the SOEP is a valuable instrument for assessing the non-coverage bias in the MC. For labour force flows the non-coverage bias of the MC appears to be only of moderate size.

JEL Classifications: C81, J69.

1. Introduction

Household panels are mainly run by academic institutions on a voluntary basis with sample sizes of about 5000 households. This holds, for example, for the Panel Study of Income Dynamics (PSID) which was started in 1968 by the US Survey Research Center, the German Socio Economic Panel (SOEP), where persons and families have been surveyed annually since 1984, and the British Household Panel Study (BHPS), run by the University of Essex since 1991.

These panels suffer from two drawbacks. First, because of the voluntary participation, there was a substantial initial nonresponse of about one-third of the sample, which was followed by non-participation in later waves, denoted

* This work is a part of the "MC-Panel" project, sponsored by the Bundesminsterium für Bildung und Forschung (BMBF) and the Deutsche Forschungsgemeinschaft (DFG). We gratefully acknowledge the helpful comments of the anonymous referee and the editor.

as panel attrition. The cumulative effect of panel attrition reduces the case numbers which aggravates the problem of how to represent rare events, such as the incidence of receiving social aid or having a very high income, in the sample. Thus the number of cases might be too small to provide conclusive evidence in the analysis at hand. Voluntary participation is often regarded as an uncontrolled source of bias.

In Germany both objections can be met by the annual micro census (MC), which can be merged to a four-year rotating panel, see (Heidenreich 2002). Its longitudinal sample size is about ten times larger than the academic national counterpart, the SOEP. Furthermore, participation in the MC is mandatory. Hence nonresponse is reduced to a minimum level.

However, a serious problem arises from the fact that residential movers are not traced in the MC. Unlike in the SOEP, this non-coverage leads to missing information for residential movers at their new home. Instead, new persons who move into the dwellings of the residential movers enter the MC sample. Such moves are not covered in the SOEP, where only moves into already existing households are recorded. Furthermore, for a reduced set of variables, the MC questionnaire asks about the previous calendar year retrospectively. Thus, for these variables the MC panel is representative and also includes residential movers. However, for other variables, there may be a potential bias due to non-coverage if reasons for mobility are correlated with the variable of interest. For example, if we are interested in changes from unemployment to employment, then a move to a different place may be prompted by a new job. In this case a weighting procedure may correct a potential non-coverage bias: The movers are dropped from the longitudinal analysis and weights are applied to the people who do not move, the so-called stayers. Typically the weighting procedure uses a logistic regression model to predict the residential mobility. The weights assigned to the stayers may be taken as the inverse of the predicted probability of being immobile, see (Clarke / Tate 2002).

Besides the differences in the treatment of residential mobility, the two surveys differ in the following respects: (1) The SOEP uses intensive oversampling of special sub-populations, namely foreigners, East Germans and immigrants, see (Haisken-DeNew / Frick 2003). In contrast, the MC uses an equal probability sample for the entire population. (2) For the time interval of our analysis, the SOEP is in wave 13 (= 1996) while our MC sample is in its first wave. (3) There are some minor differences in the design of the questionnaire. Also the time reference is not identical. Here, the MC uses a strict concept of a reference week (second week in May) whereas the SOEP refers to the moment of the interview (approximately the end of March).

There are two ways to exploit the MC and the SOEP: First, the design-based approach refers to the sampling design of the respective survey and estimates population totals and ratios by weighting with the inverse of the sampling probability, see (Särndal et al. 1992). Second, the model-based approach assumes a statistical model and independent identical distributed (iid) observations. An efficient estimate is obtained by maximizing the corresponding likelihood function. Our results refer to both approaches.

The main purpose of the paper is to decide whether the SOEP can help to detect a non-coverage bias in the MC and how far the weighting approach can correct the non-coverage bias. Furthermore, special emphasis is given to the longitudinal comparison of the MC and the SOEP results.

The paper is organized as follows: first, we display the extent of residential mobility in the SOEP and the MC. Then, we introduce our model for household mobility. With respect to the availability of the SOEP data, we then propose tests for the detection of a possible non-coverage bias. In the next step, we compare the MC and SOEP data for labour force flows. This is intended as a means to evaluate the comparability of the MC and the SOEP. Finally, we apply the weights, provided by a model for household mobility, to correct a non-coverage bias in the MC. In the last section we summarize our findings.

2. The Extent of Residential Mobility

In this section we display the frequency of residential mobility according to the MC and the SOEP. First we want to check whether the extent of mobility in the MC and the SOEP is comparable.

The MC is a rotating panel. Every year, a quarter of the sampling districts included in the sample are rotated. This means that each household remains in the sample for four years. The basis of our analysis is one rotation group of the MC, which stayed in the sample from 1996 to 1999.

Residential mobility is not an individual decision. Often mobility is clustered within households. There are, however, exceptions to this general pattern, for example, if children leave their parents' home. To take this into account we focus our analysis on "synthetic" households.[1] A synthetic household is a group of individuals living together in a physical household and having the same mobility behaviour. For example, imagine a three member household where one individual moves out between 1996 and 1999. In this case, the individual who moves out forms one synthetic household and the other two members of the original household form a second synthetic household.

Table 1 shows the extent of residential mobility according to the MC and the SOEP. The figures in row SOEP* are calculated using design weights and

1 This idea is based on (Clarke / Chambers 1998). They construct a model for household-level nonresponse.

attrition correction factors, see (Haisken-DeNew / Frick 2003). From Table 1, we can conclude that the mobility behavior is very similar. This holds for the unweighted and weighted SOEP data. According to the MC, the cumulative effect from 1996 to 1999 amounts to about 31 percent of mobile households. Here, one would expect a cumulative effect of about three times the annual rate of 13.5 percent or more because persons may change their residence more than once during three years. However, such multiple changes of residence are not recorded here.

Table 1

The Extent of Residential Mobility in %
(Unweighted and Weighted Results, Synthetic Households)

Sample		Transition 1996 – 1997	1996 – 1998	1996 – 1999
MC	Mobile	13.57	23.32	30.76
	Immobile	86.43	76.68	69.24
	Total	100.00	100.00	100.00
SOEP	Mobile	12.78	23.26	29.89
	Immobile	87.22	76.74	70.11
	Total	100.00	100.00	100.00
SOEP*	Mobile	12.27	23.69	31.04
	Immobile	87.73	76.31	68.96
	Total	100.00	100.00	100.00

Source: Authors' calculations, Data base: SOEP, MC, Waves: 1996 – 1999.

3. A Model for Household Mobility

While the extent of residential mobility appears to be comparable for MC and SOEP between 1996 and 1999, we now focus on the question of whether in both surveys comparably defined covariates exert the same impact on the probability of residential mobility. However, we do not intend to present an extensive model on residential mobility. Such an analysis can be found for the SOEP in Frick (1996). Our main interest is a suitable model for the prediction of residential mobility in the MC.[2] Based on these predictions we will calculate correction factors that compensate for the non-coverage of residential mobility, see Section 6. We run logit analyses based on the MC and the SOEP data from 1996 to 1999.

[2] An important explanatory variable of residential mobility, housing tenure, is not used here because it is not available in the MC.

Table 2

Probability of Residential (Im-)mobility over the Period 1996 – 1998 (Logit Analysis)

Variable		MC	SOEP	Diff.
Intercept		**2.0427**	**1.8933**	0.1493
		(0.0558)	(0.1652)	(0.1744)
Household size	1 person	**-1.3695**	**-1.2156**	-0.1539
		(0.0384)	(0.1093)	(0.1158)
	3 persons	**0.9030**	**0.7038**	0.1992
		(0.0436)	(0.1035)	(0.1123)
	4 persons	**1.5489**	**1.4904**	0.0585
		(0.0571)	(0.1328)	(0.1446)
	Age $\geq$ 5 persons	**1.5973**	**1.5749**	0.0224
		(0.0798)	(0.1788)	(0.1959)
Age $\leq$ 30	1 person	**-1.4083**	**-1.3737**	-0.0346
		(0.0317)	(0.0927)	(0.0980)
	$\geq$ 2 persons	**-2.2065**	**-2.1270**	-0.0795
		(0.0493)	(0.1245)	(0.1339)
Age > 60	1 person	**0.5627**	**0.8108**	-0.2481
		(0.0350)	(0.1254)	(0.1302)
	$\geq$ 2 persons	**0.2016**	**0.5604**	-0.3588
		(0.0646)	(0.2279)	(0.2369)
Household income	n / a	**-0.1625**	-0.0122	-0.1503
		(0.0445)	(0.1685)	(0.1743)
	< 2200	**0.1664**	**0.2297**	-0.0633
		(0.0379)	(0.1078)	(0.1143)
	2200 to < 3000	**0.1892**	**0.2270**	-0.0378
		(0.0403)	(0.1090)	(0.1162)
	4000 to < 5500	**-0.1256**	-0.0754	-0.0503
		(0.0411)	(0.0940)	(0.1025)
	5500 and more	**-0.2285**	-0.1675	-0.0609
		(0.0457)	(0.1101)	(0.1192)
Region	East-Germany	**-0.3310**	**-0.3885**	0.0576
		(0.0276)	(0.0786)	(0.0833)
School [No.]	secondary	0.0205	0.1196	-0.0991
		(0.0330)	(0.0872)	(0.0932)
	grammar	**0.1393**	**0.1783**	-0.0391
		(0.0248)	(0.0703)	(0.0745)
Education [No.]	vocational	**0.1759**	**0.1806**	-0.0047
		(0.0219)	(0.0609)	(0.0647)
	tertiary level	**0.2013**	0.0811	0.1202
		(0.0361)	(0.0920)	(0.0989)
Nationality	$\geq$ 1 foreigner	**-0.7313**	**-0.2868**	**-0.4445**
		(0.0429)	(0.0928)	(0.1023)
Observations		53'821	6'777	
Log Likelihood		-24'530	-3'131	
Pseudo R^2		0.1662	0.1494	

Dependent Variable: indicator of mobility coefficients for logarithm of odds ratio $P(M = 0)$ / $P(M = 1)$ (positive coefficients indicate a higher tendency for immobility) Standard deviations in paranthesis.

Source: Authors' calculations, Data base: SOEP, MC, Waves: 1996, 1998.

These analyses are based on socio-economic indicators that are observable in 1996. The variables we include in our analysis are: household income, region, household size and number of household members with respect to age group, education level, and school level. We do not include changes in socio-economic status such as labour force flows because these are not observable for mobile households in the MC. We study the effects of these items on mobility behaviour for different time intervals, in order to detect possible trends.[3]

The first column in Table 2 (MC) shows the results of the logit analyses based on the MC and the second column (SOEP) the results based on the SOEP data for the 1996 to 1998 period. The third column displays the differences between the two estimates. Significant estimates (p-value < 0.05) are indicated by bold figures. According to Table 2, age plays an important role in mobility behaviour. As expected, the number of persons over 60 years in a household increases the probability of staying at the same address. On the opposite side, the number of young persons decreases the probability of staying immobile. Similarly, the household size is a good predictor for mobility propensity. Nationality, measured by an indicator for Non-Germans in the household, has a positive impact on mobility. This is the only indicator where the MC and the SOEP differ, with higher mobility in the MC. This difference probably stems from the oversampling of special nationalities in the SOEP.[4] The remaining covariates have only a low numerical influence on the mobility behaviour. By varying the time interval, we find that the estimated slope coefficients are stable over time.[5] Only the intercept decreases with length of the time interval, which is plausible.

Finally, we obtain the remarkable result that the mobility behaviour in the SOEP and the MC, after controlling for important design variables, is equal, nationality being the only exception.

---

3 A note on the treatment of attrition: In the SOEP, all subjects participating prior to the current wave are listed on the sampling address list, regardless of whether they refuse later or not. Hence, the information about residential mobility between the years 1996 and 1997 is available for all subjects who participated in 1996. This is not the case if we consider residential mobility between 1997 and 1998 including all participants in 1996. Therefore, in our analysis we restrict ourselves to persons who participated in 1997. When we analyze mobility behaviour between 1996 and 1998, we account for a symmetric treatment of movers and stayers by dropping all subjects who did not respond in 1997. As a consequence, all movers and stayers who did not participate in 1997 are included in the analysis 1996/1997 but excluded from the analysis 1996/1998. In the MC, all movers and stayers between 1996 and 1997 are included in the analysis 1996/1998. The time interval 1996/1999 is treated in a similar way. Here the analysis is restricted to participants in the waves 1996 to 1998.

4 The SOEP started with a separate foreigner sample, the so-called Sample B, that represents the immigrant workers in 1984. Furthermore, the immigrants subsample, Sample D, overrepresents immigrants from Eastern Europe. In the MC sample there is no information about this group membership.

5 These results are available from the authors upon request.

4. The Bias Due to the Non-Coverage of Residential Movers

The bias due to the non-coverage of the residential movers cannot be derived from the MC alone. Therefore we use the SOEP, which covers residential mobility, to asses the non-coverage bias. Here we estimate the transitions between labour force states for three groups of individuals, namely the group of residential movers, the group of residential stayers and the group of the residential movers and stayers together, called "All". The difference in the transition rates between the group of all individuals and the residential stayers is an estimate of the bias due to the non-coverage of residential mobility in the MC.

In Table 3 we compare transitions between employment (E), unemployment (U) and not being in the labour force (N). The first column of Table 3 displays the transition rates for all persons, the second column for residential stayers, and the third column for residential movers. The transition rates are calculated for unweighted data and weighted data, using design weights and attrition correction factors. In order to detect possible trends in transition rates we considered the transitions between 1996/97, 1996/98 and 1996/99. A comparison of the unweighted and weighted results reveals that there are some differences between corresponding estimates. However, the difference between the column "All" and the column "Stayers" which measures the non-coverage bias is similar for unweighted and weighted data. The non-coverage bias is only slightly higher when using weighted data.

For some transitions, there are considerable differences. For example, the transition from unemployment to employment ($U \rightarrow E$) is more frequent among residential movers (48.65 percent for unweighted and 45.30 percent for weighted data) than among residential stayers (30.85 percent for unweighted and 27.17 percent for weighted data), which is plausible as the new job might have caused a change of residence. Due to the low percentage of the mover group, the resulting bias from the omission of the movers is only 1.98 percentage points for unweighted and 2.40 percentage points for weighted data. Also for the transition from not being in the labour force to employment ($N \rightarrow E$) we observe large differences between residential stayers and movers. Here, the differences also increase when the time interval becomes longer. These results indicate that there is a tendency to overestimate stability in the labour market states if only residential stayers are regarded.

To assess the non-coverage bias in the model-based approach, we perform a Hausman-test.[6] Here we check whether the difference between the estimates using only the information of stayers ($\hat{p}_{\text{stayers}}$) and the estimates using the

[6] For the design-based approach there is no similar analytical tool. Hence we were not able to present significances for the bias of the weighted results in Table 3.

Table 3

Transition between Different Types of Labour Market Status (Unweighted and Weighted Results)

Transitions from 96 to		E			U			N		
		All	Stayers	Movers	All	Stayers	Movers	All	Stayers	Movers
		unweighted results								
E	97	91.02	91.16	89.87	4.92	4.86	5.43	4.05	3.97	4.70
	98	87.82	88.03	87.06	**6.32**	**6.04**	**7.36**	5.86	5.93	5.58
	99	**87.01**	**86.37**	**88.63**	6.04	6.30	5.33	6.96	7.33	6.04
U	97	**32.83**	**30.85**	**48.65**	**48.39**	**49.83**	**36.94**	18.78	19.32	14.41
	98	**34.92**	**31.79**	**47.09**	40.13	41.20	35.98	24.95	27.01	16.93
	99	**41.37**	**37.46**	**50.99**	28.91	29.10	28.46	29.71	33.44	20.55
N	97	**12.74**	**11.64**	**25.65**	5.48	4.97	11.52	81.77	83.39	62.83
	98	**19.66**	**16.07**	**38.04**	**5.09**	**4.40**	**8.65**	75.25	79.54	53.31
	99	**25.89**	**21.13**	**42.01**	**4.53**	**3.71**	**7.31**	69.58	75.15	50.68
		weighted results								
E	97	90.65	90.35	92.83	4.53	4.70	3.30	4.82	4.95	3.88
	98	87.49	87.55	87.32	5.97	5.44	7.65	6.54	7.01	5.03
	99	86.89	85.43	90.08	5.42	5.90	4.35	7.70	8.67	5.57
U	97	29.57	27.17	45.30	48.55	49.26	43.94	21.88	23.57	10.77
	98	33.28	27.14	54.91	39.10	41.84	29.47	27.61	31.02	15.63
	99	37.82	30.95	53.47	27.44	28.98	23.94	34.73	40.07	22.59
N	97	12.74	11.35	28.50	4.12	3.73	8.57	83.14	84.93	62.93
	98	17.16	12.36	39.18	4.35	3.92	6.30	78.49	83.72	54.52
	99	22.92	17.51	40.43	4.07	3.09	7.24	73.00	79.40	52.33

Source: Authors' calculations, Data base: SOEP, Waves: 1996-1999.

information of movers plus stayers ($\hat{p}_{\text{all}}$) is significant. The estimator $\hat{p}_{\text{all}}$ is efficient because it is the ML-estimate based on all available information. The estimator $\hat{p}_{\text{stayers}}$ which is based only on the subsample of stayers is consistent under the null-hypothesis of no bias due to residential moves.[7] Significant differences (p-value < 0.05) are indicated by bold figures.[8] According to the Hausman test, we find that especially for the transitions from not being in

Table 4

Logit analysis: Transition from Unemployment to Employment. Time Interval 1996 to 1998

Variable		All	Stayers	Diff.
Intercept		**-0.5989**	**-0.6084**	0.0095
		(0.0003)	(0.0009)	(0.9058)
Age	≤ 30	**1.2521**	**1.3920**	-0.1399
		(0.0001)	(0.0001)	(0.1915)
Education	without vocational	-0.2351	-0.1707	-0.0644
		(0.2272)	(0.4530)	(0.5840)
	tertiary level	0.4872	0.3447	0.1425
		(0.0931)	(0.3114)	(0.4231)
Region	West-Germany	-0.1240	-0.1467	0.0277
		(0.4651)	(0.4592)	(0.8246)
School	without exam	0.0785	-0.0689	0.1474
		(0.7860)	(0.8384)	(0.3990)
	grammar	0.2165	0.1205	0.0960
		(0.4150)	(0.6972)	(0.5463)
Sex	Male	**0.3841**	0.2558	0.1283
		(0.0135)	(0.1566)	(0.1617)
Duration of unemployment	> 1 year	**-1.2364**	**-1.3065**	0.0701
		(0.0001)	(0.0001)	(0.4479)
Observations		899	715	
Log Likelihood		-503	-383	
Pseudo R^2		0.1570	0.1647	

p-values in paranthesis.

Source: Authors' calculations, Data base: SOEP, Waves: 1996, 1998.

7 The behaviour of the efficient and the consistent estimator under the alternative is exchanged here. In econometric application, the efficient estimator becomes inconsistent while the consistent estimator remains consistent under the alternative. As the Hausman test is evaluated under the null hypothesis, this change is irrelevant here.

8 We tested the hypothesis that the bias of the transitions to U and E is equal to zero. This implies also a zero bias for the transition to the remaining category N.

labour force / unemployment to employment (N→E and U→E), the observed differences in Table 3 are significant and of remarkable size. In what follows, we use control variables for the labour force transitions. We present the results of logit analyses for both groups, namely the group "All" of stayers plus movers, and then the group of stayers only.[9] The idea of this analysis is that the above differences in transition rates of the two groups are connected with differences in socio-demographic characteristics. In the context of missingness definitions, this would be equivalent to the missing at random (MAR) assumption if observed control variables are used. We model the transition from being unemployed to employment. We use explanatory variables that are considered to affect the employment status: age, sex, region, school / education level, duration of unemployment.[10] Let β denote the slope parameter of a Logit model for the successful transition between different labour force states. If $\hat{\beta}_{\text{all}} = \hat{\beta}_{\text{stayers}}$ holds, this could be seen as indicating the case of MAR.

The estimates of the model parameters are displayed in Table 4. The first column (All) in Table 4 shows the results obtained using movers as well as stayers while the second column (stayers) shows the results obtained using the stayers only. The third column (Diff) in Table 4 displays the differences of the estimated coefficients. Significant differences (p-value < 0.05) are indicated by bold figures. We observe that the younger persons have a significantly higher probability of making transition from unemployment to employment in comparison to persons older than 30 years. As expected, the long-term unemployed are less likely to find employment. The rest of the covariates yield only a small influence on making the transition from unemployment to employment. The above interpretation holds for both samples.

To determine whether the effects of covariates differ by mobility, we apply a Hausman test for the differences in the coefficients. According to the Hausman test, there are no significant differences between the two samples. Thus, the results from Table 4 suggest that by conditioning on observed control variables the non-coverage bias in labour force transitions can be reduced if not removed. This holds also for longer time spans.[11]

---

9 This analysis is not done for the movers group because of different population sizes of movers and stayers. In a direct comparison of the mover and stayer groups, the non-coverage bias would be overestimated as the mover group is much smaller than the stayer group.

10 We use only variables that are available in the MC. Thus, we do not include changes of socio-economic states because these are not observable for mobile households in the MC.

11 By varying the time interval we reached similar results, i.e. no significant differences between estimated coefficients.

5. Comparison of Longitudinal Analysis: MC vs SOEP

In this section we compare the transitions between labor force states for the stayers based on the MC and the SOEP. The motivation for this comparison is to test whether the MC and the SOEP yield similar labour market mobility figures. If this is the case, then the methods of bias control via the SOEP appear to be more reliable, as the two surveys' measurements produce comparable results. In Table 5, we displayed the unweighted (column SOEP) and weighted (column SOEP*) results for the SOEP. We also performed a test[12] to check whether the differences between the estimates based on the MC and the SOEP / SOEP* are significant. For the variance estimation of the SOEP* we used the random group approach, see (Haisken-DeNew / Frick 2003). Significant differences (p-value < 0.05) between MC and SOEP are indicated by bold figures. The results in Table 5 reveal that differences between corresponding estimates of MC and SOEP are small, the only exceptions being transitions from unemployment to unemployment / not being in labour force for the time interval 96 / 99. However, some of these small differences are significant[13] and stable over time.

A comparison of the MC and the design-weighted SOEP estimates reveals an increase of the differences for transitions starting from unemployment and not being in the labour force. However, transitions starting from employment fit nicely the MC figures.

There are several possible reasons for these differences. First, the differences for the unweighted SOEP could be due to different sampling designs and nonresponse. But the results for the SOEP* indicate that this is not true.[14]

Second, the phrasing of the two questionnaires with respect to labour market definitions is similar, but not identical for the two surveys. Here, we found that these estimates are very sensitive to the underlying definition. Thus, only minor deviations in item definition could lead to remarkable differences between corresponding estimates, see (Rendtel et al. 2004).

Third, the fieldwork periods for the two surveys do not coincide fully, taking place in March for SOEP and May for MC. Thus, seasonal effects could

12 We performed a two group test, which is based on the fact that two samples are independent from each other and on the following test statistic $t = \frac{(\hat{p}_{\text{MZ}} - \hat{p}_{\text{SOEP}})}{\sqrt{\hat{\sigma}^2_{\text{MZ}} + \hat{\sigma}^2_{\text{SOEP}}}} \sim \mathcal{N}(0,1)$.

13 This could be due to the fact that the case numbers in MC are very high, resulting in small estimated standard errors.

14 The same analysis was repeated by using only data for a subsample including only West Germans. For this subsample, the SOEP and the MC designs are both an equal probability sample. However, the differences between the two surveys did not decrease, indicating that design matters are not responsible for these differences.

Table 5

Longitudinal Results for Transition between Labor States for Immobile Persons

Transition from 96 to		E			U			N		
		MC	SOEP	SOEP*	MC	SOEP	SOEP*	MC	SOEP	SOEP*
E	97	90.59	91.16	90.35	4.21	**4.86**	4.70	5.20	**3.97**	4.95
	98	87.57	88.03	87.55	4.95	**6.04**	5.44	7.48	**5.93**	7.01
	99	85.32	86.37	85.43	5.16	**6.30**	5.90	9.52	7.33	8.67
U	97	29.29	30.85	27.17	52.20	49.83	49.26	18.51	19.32	**23.57**
	98	32.53	31.79	27.14	42.39	41.20	41.84	25.08	27.01	**31.02**
	99	35.89	37.46	**30.95**	34.60	**29.10**	28.98	29.51	**33.44**	**40.07**
N	97	13.11	**11.64**	11.35	3.77	**4.97**	3.73	83.12	83.39	84.93
	98	17.69	16.07	**12.36**	3.41	4.40	3.92	78.90	79.54	**83.72**
	99	22.14	21.13	**17.51**	2.98	3.71	3.09	74.89	75.15	**79.40**

Source: Authors' calculations, Data base: SOEP, MC, Waves: 1996-1999.

affect the estimates. This can be seen from Table 6, where we compare cross-sectional estimates of labour market states. The results in Table 6 reveal that SOEP / SOEP* shows more people unemployed and that MC shows more people not in the labour force. These results suggest that the differences in Table 6 may be partially due to the annual seasonal pattern in the unemployment rate and consequently that the differences in Table 5 result from these different levels of unemployment.

Table 6

Cross-Sectional Results for Labor States

	E				U				N			
	1996	1997	1998	1999	1996	1997	1998	1999	1996	1997	1998	1999
MC	65.03	64.38	64.44	65.43	6.89	7.62	7.48	7.09	28.07	27.99	28.07	27.49
SOEP	68.33	67.68	67.67	69.81	8.86	9.13	9.26	7.75	22.81	23.19	23.07	22.45
SOEP*	67.23	66.36	66.60	68.82	8.19	8.40	8.66	7.23	24.57	25.24	24.74	23.95

Source: Authors' calculations, Data base: SOEP, MC, Waves: 1996–1999.

6. Bias Correction by Weights

Now we turn to the question of whether weighting by inverse mobility rates reduces the bias of the estimated labour force flows. Therefore, we used the reciprocal probabilities obtained from the household mobility model in Section 3. More precisely, we assign to each member of an immobile household the weight $\hat{\mathrm{P}}(\mathrm{M} = 0|X)^{-1}$ and to each member of a mobile household the weight zero. In Table 7 we display the relative bias (in %) of the transition rate estimates between labour force states, estimated with the above weights and without weights. The bias is estimated for unweighted and design-weighted SOEP data, as in Section 4. We obtain two bias estimates: the uncorrected bias is obtained by using no weights from the mobility model for the stayers while the corrected bias is obtained by using the above weights for the stayers. The weighting leads to improved estimates if the corrected bias is smaller than the uncorrected bias. This applies if the bias is large, for example in the case of the transition from not being in the labour force to employment (N $\rightarrow$ E). Here the bias is reduced to some extent. These cases are indicated in Table 7 by bold figures. This holds for unweighted as well as weighted data. Similar results were shown by (Neukirch 2002).

Table 7

Relative Biases of Transitions

Transition	1996 – 1997		1996 – 1998		1996 – 1999	
	uncorrected	corrected	uncorrected	corrected	uncorrected	corrected
	unweighted results					
EE	0.13	0.34	0.28	0.75	**0.83**	**0.05**
EU	1.06	2.24	4.93	7.28	**4.54**	**0.13**
UE	**6.00**	**4.81**	**9.31**	**6.60**	**10.09**	**3.13**
UU	**2.95**	**2.87**	**2.82**	**2.77**	0.64	1.04
NE	**8.90**	**5.82**	**17.44**	**12.36**	**18.68**	**8.85**
NU	**8.21**	**6.91**	**13.38**	**9.33**	**15.91**	**11.96**
	weighted results					
EE	**0.32**	**0.13**	0.14	0.60	**1.74**	**0.58**
EU	**3.88**	**3.02**	9.44	12.21	**8.32**	**2.57**
UE	**8.03**	**5.18**	**18.65**	**15.49**	**18.96**	**6.77**
UU	1.38	2.28	7.11	10.96	6.29	7.29
NE	**11.26**	**6.97**	**27.51**	**21.55**	**23.48**	**11.05**
NU	**8.67**	**6.80**	**10.05**	**7.76**	**23.56**	**20.69**

Relative bias in %.
Source: Authors' calculations, Data base: SOEP, Waves: 1996-1999.

7. Conclusion

Although the MC was originally designed for cross-sectional purposes, it can be used for longitudinal analysis of up to four annual measurements. In this paper we explored the effect of non-coverage due to residential mobility on labour force flows. We estimated this non-coverage bias from the SOEP data and applied a weighting approach to correct a potential non-coverage bias.

The results show that there is a non-coverage bias with respect to some transitions. Especially for transitions from unemployment to employment as well as from not being in the labour force to employment there exists a correlation between labour market mobility and regional mobility. The use of some control variables, for example age, reduces this non-coverage bias. This can also be verified by a mobility model. Here, the age has the biggest impact on the mobility.

In order to evaluate the comparability of the SOEP and the MC results we compared the labour force flows and logit coefficients for household and labour force mobility. For labour force mobility, we found that corresponding estimates do not coincide fully. However, the differences are in general small and may also be related to minor deviations in item definitions and different fieldwork periods. For household mobility, a good match was obtained despite the fact that this comparison is based on the 13th to 16th waves of the SOEP while the MC is evaluated after its first four waves.

Finally, we used the weights from our mobility model to reduce the non-coverage bias in labour force flows. The results suggest that weights produce flows estimates that are less biased. The longer the time interval, the higher the bias reduction due to the application of weights. Overall, our results suggest that for labour force flows, the MC can be used as a panel. For other variables the same methodology as presented here should be applied.

References

Clarke, P. S. / *Chambers,* R. L. (1998): Estimating Labour Force Gross Flows From Surveys Subject to Household-level Nonignorable Nonresponse, Survey Methodology 24, 123 – 129.

Clarke, P. S. / *Tate,* P. F. (2002): An Application of Non-Ignorable Non-Response Models for Gross Flows Estimation in the British Labour Force Survey, Australian & New Zealand Journal of Statistics 44, 413 – 425.

Frick, J. (1996): Lebenslagen im Wandel: Determinanten kleinräumlicher Mobilität in Westdeutschland. In: Sozio-ökonomische Daten und Analysen für die Bundesrepublik Deutschland, Band 9 (Dissertation), Frankfurt / New York.

Haisken-DeNew, J. / *Frick,* J. R. (2003): Desktop Companion to the German Socio-Economic Panel Study (GSOEP), Version 7.0 – Update to Wave 19, DIW Berlin.

Heidenreich, H.-J. (2002): Längsschnittdaten aus dem Mikrozensus. Basis für neue Analysemöglichkeiten. [Longitudinal Data on the Basis of the German Microcensus. A new data base for analysis], Allgemeines Statistisches Archiv 86, 213 – 231.

Neukirch, T. (2002): Nonignorable attrition and selectivity biases in the Finnish subsample of the ECHP, CHINTEX Working Paper #5, URL: http://www.destatis.de/chintex/download/paper5.pdf

Rendtel, U. et al. (2004): Report on quality of income data, CHINTEX Working Paper, 21, URL: http://www.destatis.de/chintex/download/paper21.pdf

Särndal, C.-E. / *Swensson* B. / *Wretman,* J. (1992): Model assisted Survey Sampling, New York.