

## **A Representative Matched Cross-section Survey for Austria - Measuring Worker Flow Dynamics with the Austrian Labour Force Survey**

Schoiswohl, Florian; Wüger, Michael

Published: 01/03/2016

### *Document Version*

Publisher's PDF, also known as Version of record

[Link to publication](#)

### *Citation for published version (APA):*

Schoiswohl, F., & Wüger, M. (2016). *A Representative Matched Cross-section Survey for Austria - Measuring Worker Flow Dynamics with the Austrian Labour Force Survey*. (Department of Economics Working Paper Series; No. 223). WU Vienna University of Economics and Business.

Department of Economics  
Working Paper No. 223

# A Representative Matched Cross-section Survey for Austria - Measuring Worker Flow Dynamics with the Austrian Labour Force Survey

Florian Schoiswohl  
Michael Wüger

March 2016



# A Representative Matched Cross-section Survey for Austria

## Measuring Worker Flow Dynamics with the Austrian Labour Force Survey

Florian Schoiswohl<sup>†</sup>

Michael Wüger<sup>‡</sup>

February, 1st 2016 , Vienna

### Abstract

While worker flow analysis has grown in importance in many countries, Austria still lacks a specific longitudinal dataset as a prerequisite to perform similar analyses. For this reason, this article provides a coherent procedure to construct a longitudinal dataset based on the rotational panel structure of the Austrian quarterly LFS from 2004 to 2014. The procedure, which is available for researcher, is grounded on the discussion of several related and important issues inherent in constructing this sort of longitudinal data: First, it deals with the construction of the quarterly-matched dataset and the quality-of-measurement of several labour market variables. Second, the paper analyses non-response as a sample selection process, and shows that the selected (quarterly-matched) dataset causes biased estimates of worker flows. Third, the article proposes an iterative raking procedure to obtain survey weights as a bias-correcting device for any future analysis. Based on these adjustments, we present unbiased time-series of worker flows and transition rates, and conclude that the employment-unemployment margin is highly sensitive to economic shocks and that the Austrian labour market is additionally shaped by large movements within the participation margin.

**JEL classification:** C81; J21; J63

**Keywords:** Matched cross-sections; Sample selection; Worker flows; Unemployment dynamics; Austrian labour market

---

<sup>†</sup>Institute for Labor Economics, [florian.schoiswohl@wu.ac.at](mailto:florian.schoiswohl@wu.ac.at)

<sup>‡</sup>Institute for Labor Economics, [michael.wueger@wu.ac.at](mailto:michael.wueger@wu.ac.at)

# 1 Introduction

Worker flow analysis has grown in importance in many countries and has emerged to a key framework in analysing labour market dynamics (Petrongolo & Pissarides 2008, Elsby et al. 2009, Fujita & Ramey 2009, Shimer 2012, Elsby et al. 2013, 2015). Most of the papers in this field of literature base their estimation of worker flows and transition rates on appropriate longitudinal labour force surveys which are often constructed by matching cross-section datasets. Unfortunately, Austria lacks such a specific longitudinal dataset which might be one important reason why Austrian labour market dynamics have therefore not been analysed so far<sup>1</sup>. This is surprising given that some recent labour market developments (e.g. the positive co-movement of the labour force participation and the unemployment rate since 2011) should be properly analysed by focusing on worker flows and the transition behaviour of individuals.

Against this background, we are the first to employ the rotational panel structure of the Austrian quarterly LFS from 2004-2014 to construct a longitudinal dataset which is appropriate for the analysis of worker flows<sup>2</sup>. For this purpose, we provide a coherent procedure (including codes) which deals with several important issues arising when such a dataset is constructed. Thereby, our methodological discussion rest upon three major parts.

First, we are matching individuals between two consecutive quarters to identify worker flows on a quarterly frequency<sup>3</sup>. For the US counterpart of the LFS (CPS), Abowd & Zellner (1985) and Poterba & Summers (1986) have documented significant misreporting in labour market variables which, in turn, leads to biased worker flow estimates. Therefore, we devote some effort to check the quality-of-measurement with respect to the labour status variable.

Second, due to non-response in the survey period, matching cross-section across quarters leads to an incomplete panel dataset. This may result in a sample selection bias when estimating worker flows. Biased estimates would not be the case if one can assume that

---

<sup>1</sup> Notable exceptions are Hofer et al. (2001) and Stiglbauer et al. (2003) who have used the social security dataset for their analyses.

<sup>2</sup> Note, that so far only Eurostat has undertaken some efforts in exploiting the longitudinal component of the Austrian LFS by providing first experimental series on worker flows in Europe (Eurostat 2015). However, although its procedure is similar to ours, it provides only series for the population aged 15-74 years, only stratified by sex and only since 2010Q2. Furthermore, the published meta-data neither includes a thorough discussion of the data quality, nor a possibility to access the longitudinal microdata for reproducing the published series.

<sup>3</sup> We are aware of the possibility to construct annual or annualized flow measures considering either matching between all five available interviews, or only the first and the last interview. We follow the international trend and provide a two-quarter matched dataset, and encourage researchers to adapt our procedures to any other form of matched dataset.

the sample selection is independent of the behavioural relationship of interest (e.g. labour market transitions). However, for our specific kind of data, this assumption is most likely violated, as the non-response behaviour of individuals certainly depends on labour market related variables<sup>4</sup>. The existence of such bias has already been studied by Hausman & Wise (1979) and van den Berg et al. (1994) in case of linear and duration models, and by Peracchi & Welch (1995) and Jimenez-Martin & Peracchi (2002) specifically for labour market transitions. We therefore follow this literature and analyse the determinants of the sample selection process and investigate the nature of bias in worker flow estimates that may arise due to this selection. Third, due to the existence of a sample selection bias in worker flow estimates, we propose an intuitive re-weighting procedure to obtain survey weights as a bias-correcting device for any future analysis<sup>5</sup>.

Finally, the adjusted dataset is then used to provide a short overview of worker flows and transitional behaviour of individuals in Austria. Our results suggest that Austria is characterized by relatively large worker flows and transition rates compared to other European countries and also to those of the US economy. Furthermore, we find that the employment-unemployment margin is highly sensitive to economic shocks and that the Austrian labour market is shaped by sizeable movements within the participation margin, especially between unemployment and inactivity in the aftermath of the Financial Crisis.

The paper is structured in the following way: section 2.1 starts with several basic definitions concerning the labour market system, section 2.2 and 2.3 deal with the construction of the matched dataset and the measurement issues of several important variables. Section 2.4 proceeds with the analysis of matching probabilities and its determinants. Sections 3 and 4 discuss the potential sample selection bias due to matching and propose a re-weighting procedure for bias removal. Section 5 presents the estimated worker flows and transition rates, and gives a first impression of worker flow dynamics on the Austrian

---

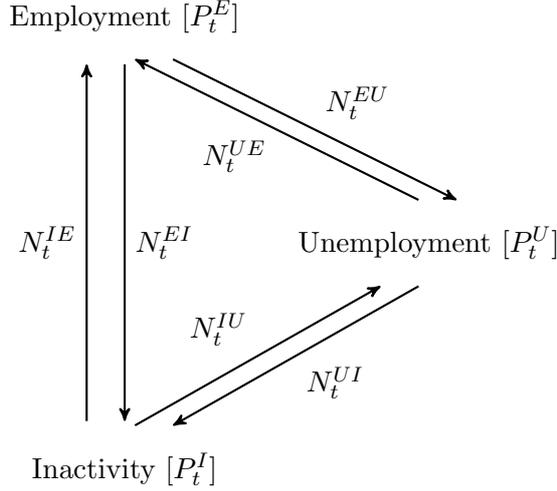
<sup>4</sup> Respondents are sampled based on a population of addresses without any obligation to follow individuals if they move away from a sampled address. As a result, all determinants of migration such as unemployment are also related to the non-response behaviour.

<sup>5</sup> In such analyses, another important issue is typically considered in the literature, but is not covered here: the influence of the frequency-of-measurement of labour market states of individuals on the size of worker flow estimates; a phenomenon labelled as time-aggregation bias. Note that a low frequency (i.e. the time between interviews) may lead to an underestimation of worker flows as many transitions remain undetected between the two survey interviews. Shimer (2005, 2012) and Gomes (2010) have documented such a bias for US worker flow estimates. This literature argues that a pro-cyclical time-aggregation bias (e.g. in recessions less short-term unemployed) renders the rate of separation from employment acyclical. This conclusion and the way of correction of the time-aggregation bias is criticized by Fujita & Ramey (2009), Elsby et al. (2009, 2010, 2013). Nordmeier (2012) bases her analysis on German administrative data and concludes that the adjustment approach proposed by Shimer (2012) does not sufficiently correct for the bias in German worker flow series. Although we provide these procedures implemented in our code-files, we encourage the readers to critically review this literature and use these files until a more comprehensive analysis of the time-aggregation issue for Austrian LFS data will be available.

labour market. And Finally, section 6 concludes the paper.

## 2 Data and Matching

### 2.1 Basic definitions



**Figure 1:** Schema of working-age population: three state labour market and gross worker flows

As depicted in Figure 1, labour market activity can be described by distinct states in which an individual can be, such as unemployment  $U$ , employment  $E$  and inactivity  $I$  (also known as non-participation in the labour force or out-of-labour force). Abstracting for now from flows between the working-age population and other states (such as reaching age 15, death or dropping out of the working age population at age 65), all stock variables  $P_t^j$  of state  $j$  are the result of gross worker flows  $N_t^{ij}$  between the state  $i$  in period  $t - 1$  and  $j$  of the subsequent period  $t$ . Additionally, we can define the transition rate  $\lambda_t^{ij}$  as the probability that an individual makes a transition from state  $i$  to  $j$ <sup>6</sup>. Let  $Y_t$  and  $Y_{t-1}$  denote the indicators of the labour market states at two consecutive points in time given,

<sup>6</sup> We do not formally differentiate between the rate and the probability of transiting from state  $i$  to  $j$  throughout the paper but always acknowledge  $\lambda_t^{ij}$  as the discrete time *rate*. However, we could transform the rate to a probability  $\Lambda_t^{ij}$  by assuming a specific event-process, e.g. a Poisson process of arrival of a specific transition shock  $ij$  with the constant rate  $\lambda_t^{ij}$ . Hence, the respective (cumulative) probability to transit at *any* point-in-time within the time interval  $\Delta t = 1$  would be  $\Lambda_t^{ij} = 1 - \exp[-\lambda_t^{ij}]$ . However, after applying such a procedure we conclude that this would not significantly change the results presented here.

then

$$\lambda_t^{ij} = P[Y_{t-1} = i, Y_t = j] = \frac{N_t^{ij}}{P_{t-1}^i} = \frac{N_t^{ij}}{\sum_j N_t^{ij}} \quad \text{with: } i, j \in \{E, U, I\}. \quad (1)$$

The worker flows  $N_t^{ij}$  and the respective transition rates can be estimated from a dataset by counting all individuals with labour market states  $i$  and  $j$  in the respective periods. Typically, counts are weighted by current period  $t$  dataset weights which enables to infer population estimates from estimates based on a sub-sample of that population (i.e. survey sample). Therefore, estimation of correct population figures for flows and transition rates requires a longitudinal dataset to be representative for the population and accurate in measuring labour market states (over time). In terms of the Austrian LFS, we will deal with both requirements in the next sections.

## 2.2 Data

The official quarterly micro-dataset of the Mikrozensus is representative for the whole population of Austria and contains the AKE/LFS (Arbeitskräfteerhebung); a specific section designed for labour market related purposes. In this study, the dataset is provided by Statistik Austria confidentially and comprises all cross-sections from 2004Q1 till 2014Q4<sup>7</sup>. Each quarterly cross-section contains the records of around 22,500 households (or approx. 45,000 individuals) who are sampled and organized in a rotational random-sample scheme<sup>8</sup>. Each sampled household is surveyed for five consecutive quarters (waves); hence each individual is observed for a whole-year period. In each quarter, one fifth of the sample drops out (out-rotation), while it is replaced by a new rotation group (in-rotation). Assuming equal-sized rotation groups and no reduction in sample sizes over interviews other than in- or out-rotation, we would be able to match eighty percent (four fifths) of the cross-section samples between two consecutive quarters. As a result, matching gives us a short panel to estimate quarterly worker flows and transition rates. In this paper, we concentrate on quarterly transitions only and ignore the possibility to construct annual transitions as well<sup>9</sup>.

Using the LFS for such a purpose has several important advantages. It has a large sample

<sup>7</sup> In principle, Austrian LFS data starting from 1994 could be used to analyse worker flows. However, the survey change in 2003/2004 led to a major structural break in the data and can not be merged without further efforts. Thus, we concentrate on the consistent data from 2004 to 2014.

<sup>8</sup> The sampling scheme randomly selects addresses of households' residences and not specific individuals.

<sup>9</sup> This would involve to match five cross-section together, by keeping only those individuals with answers in all five waves. Although not explicitly presented here, we encourage readers to adjust the provided code for this purpose.

size with a huge set of variables that contain information on the characteristics of each household-member. Moreover, it builds the basis for official and international labour statistics. For each cross-section, the survey is representative for the whole population and follows European and international definitions, such as the definitions of the labour market states according to ILO (1982). Hence it ensures comparability across countries.

The matching process using the Austrian LFS is greatly simplified as Statistik Austria provides a unique household as well as an unique individual ID variable (Statistik Austria 2015). As unique IDs are allocated to new persons and/or new households it is not necessary to use several individual and household characteristics for matching individual records<sup>10</sup>.

Despite the existence of the high-quality ID-variable, several records can not be matched between two consecutive quarters for other reasons than in- or out-rotation. Typically, this is caused by non-response in one or more survey interviews which renders the matched dataset to a selected sub-sample of the whole population sample. The consequences of non-response on *cross-section* estimates have been evaluated by Mitterndorfer et al. (2007), Gumprecht & Oismüller (2013), Meraner et al. (2015). They concluded that cross-section estimates suffer from some selection bias which leads the statistical office to change the internal processing. According to the early work of Hausman & Wise (1979), longitudinal variables may still suffer from the bias, although cross-section variables are unbiased. Furthermore, the measurement of variables may be inconsistent over time because of a change in the measurement-errors process between two consecutive periods, or changes in the interview-procedure and imputation methods (Watson & Wooden 2009). This article specifically concentrates on the consequences of these issues and on the solutions in a panel context by following the work of Peracchi & Welch (1995) and Jimenez-Martin & Peracchi (2002).

However, before we draw the readers attention to the next subsections, it is worth noting that there exist potential alternative datasets for our purpose of estimating worker flows in the Austrian labour market. To our knowledge, Stiglbauer (2003) has been the only work for Austria dealing with worker flows, while Hofer et al. (2001), Stiglbauer et al. (2003), and Aumayr (2010) concentrate on job flows rather than on worker flows<sup>11</sup>. In contrast to our work, all these empirical studies use official employment records from the Austrian social security office (Hauptverband der österreichischen Sozialversicherungsträger).

---

<sup>10</sup> For an example of matching procedures without unique ID-variables, see Peracchi & Welch (1995) or Shimer (2012).

<sup>11</sup> Job flows are measured on the firm side mostly by using employer-employee datasets. Worker flows are measured in the individuals' side using individual datasets, such as the LFS. Job flows typically underestimate worker flows, as if jobs are neither created nor destroyed within a firm, workers might very well flow in/out of the firm.

This dataset is an employer-employee database and covers all officially registered employees in the Austrian private-sector. As the dataset also contains labour market histories of individuals, it would be possible to conduct worker flow analyses. The advantages of this administrative dataset compared to the LFS are its huge size and high frequency (daily/monthly). However, the social-security-dataset suffers from several important drawbacks: it (i) serves primarily for the purpose of social security administration; hence, the identification of labour market states is not consistent with ILO-standards which heavily affects the measurement of worker flows, and (ii) it further lacks important individual characteristics, such as working hours, immigrant status or household characteristics.

Beside that, another database is given by the EU-SILC panel for Austria. Flek & Mysíková (2012) has utilized the questions of monthly activity in the survey to estimate labour market transitions for several Eastern European countries<sup>12</sup>. They argue that worker flow analyses using EU-SILC are comparable across countries and may help to observe transitions at a higher frequency than with other datasets. However, EU-SILC's monthly activity is measured through self-reporting based on a year long-term retrospective of individuals. It is likely that this framework is afflicted by reporting errors or specific reporting differences between countries instead of properly measuring cross-country differences in worker flow behaviour. Additionally, sample sizes of the EU-SILC dataset are typically much smaller than in the LFS, and it is also not free from non-response issues. For all these reasons it is necessary to establish the Austrian Labour Force Survey as a major source for analysing labour market dynamics, because it is the only dataset which ensures consistency with international measurement standards as well as providing relatively large sample sizes paired with sufficient background information on the individual level.

### 2.3 Measurement error

After motivating the need to use the LFS, we turn our attention to the longitudinal consistency of survey variables over consecutive survey-waves. In general, survey data is always prone to measurement errors in several phases of the survey, but - although typically unknown - theoretical propositions and basic descriptives on the interview process may help to gain knowledge over the potential severity of misreporting in the data. That is, several misreporting processes either depend on the nature of the variables, or on the state or wave of the survey. We will shed some light on measurement in this section by using the original cross-section datasets of the Austrian LFS without explicitly matching individuals over waves. We restrict ourselves to the working-age population (aged 15-64

---

<sup>12</sup> Bachmann et al. (2015) have also used EU-SILC for their cross-country analysis. However, they ignore the monthly information on labour market status and rely only on the annual panel information. This most likely leads to an underestimation of (short-term) flows.

Sex in $t - 1$	Sex in $t$					
	15-29		30-49		50-64	
	Men	Women	Men	Women	Men	Women
Men	99.83	0.18	99.86	0.13	99.88	0.11
Women	0.17	99.82	0.14	99.87	0.12	99.89

*Remarks: Restricted to working-age (15-64 years) only. 1,376,303 records in total of 2004Q1-2014Q4*

**Table 1:** Percentage change in sex by age

years) only. However, at some point we compare the time-consistency of several variables which requires to match individuals over time, but concentrate on potential measurement errors only.

Typically, measurement errors in survey data are studied by using re-interviews to compare the answers of the same individuals on two different questionnaires, especially in cases in which little information is available on specific error processes. Abowd & Zellner (1985), Poterba & Summers (1986), and Köhne-Finster & Lingnau (2008) use such a procedure for evaluating errors in labour force status variables. For the Austrian LFS no re-interview data is available, but we may gain some knowledge over the nature of errors in labour force status using the repeated observation of characteristics in the matched dataset. Thus, in terms of measurement issues only time-invariant characteristics can be studied properly. Nevertheless inspecting these characteristics can also shed light on potential errors in time-varying measures, such as the labour market states.

Concentrating on sex as a major demographic variable at first, Table 1 reports the probability that an individual reports different sexes in two consecutive interviews for three age groups. It can be clearly seen that (i) misreporting in sex is a negligible phenomenon in general, and (ii) it is not dependent on age<sup>13</sup>. Furthermore, in contrary to other surveys, our matching does not involve a matching error due to the usage of the unique personal identifier. Thus, the errors shown in the table are purely reporting/coding errors. Excluding erroneous records from the final dataset will not cause any spurious selection along the sex/age-dimension.

Misreporting in age may also be an important evidence of errors in demographic variables. Age is derived from the date of birth and is potentially erroneous due to a misreported date and/or calculation mistakes<sup>14</sup>. Table 2 gives the probability that an individual has a non-constant birth-year between two consecutive interviews. We see that misreporting

<sup>13</sup> We also checked the results by citizenship. Although misreporting is little higher for citizen from the 13 new member states and Extra-EU countries, the issue remains negligible.

<sup>14</sup> We concentrate on the *year* of birth only as the month of birth is not available throughout the whole sample period.

Change in birthyear	Age		
	15-29	30-49	50-64
No	99.59	99.65	99.72
Yes	0.41	0.35	0.28

*Remarks: Restricted to working-age (15-64 years) only. 1,376,303 records in total of 2004Q1-2014Q4*

**Table 2:** Percentage change of birth-year by age

the year of birth is also a negligible error source and that there is only a weak dependence on the age. However, a consistently reported year-of-birth does not imply a consistently reported age over waves. As a result, we are not only dropping all records of individuals with errors in the year-of-birth, but also those individuals who show unreasonable changes in age over waves<sup>15</sup>.

To sum up, dropping all cases with a changing sex, and an unreasonable age or year of birth reported leads to a total loss of 19,341 observations (7,581 people). Compared to 1,376,303 records (330,756 individuals) of all working-age individuals between 2004Q1-2014Q4 in our dataset (after deletion), this loss is negligible and will not significantly influence the representativeness of the dataset.

Beside errors in measurement of important demographic characteristics, variables measuring the current labour market state of the respondent are of particular interest, too. Abowd & Zellner (1985) and Poterba & Summers (1986) report misreported labour market states for the CPS. Their analyses give rise to spurious transitions between labour market states and document an over-reporting for transitions along the unemployment and inactivity (U-I) margin. This results in an under-reporting of the duration and an over-estimation of the dynamics of unemployment. Both studies report classification error rates for the U-I margin to lay between 10 and 18 percent.

Although we can not employ a re-interview study to analyse the reporting behaviour of several respondents, we may take a short look on the U-I margin. The working hypothesis is based on the assumption that the differentiation between unemployment and inactivity is very vague for some respondents when answering the survey questions. This is due to a job-search effort very close to the detection point of statistical offices (e.g. due to the definition of *active* search) which might lead to (wrongly) classify an individual as inactive rather than unemployed (searching)<sup>16</sup>. The result are spurious switches between both states caused by only minor changes in the job-search activity. Many switches between U and I

<sup>15</sup> Unreasonable changes in age are defined as changes where age decreasing, or where it increases by more than one year. Additionally, the number of changes in age (birthdays) must not be greater than one.

<sup>16</sup> See Frijters & van der Klaauw (2006) for a discussion of the U-I margin

may therefore be an indicator of misclassification, while a steady reporting behaviour in terms of inactivity may give an indication for a clear differentiation between unemployment and inactivity.

Taking all records of sampled individuals, one may observe up to five quarterly labour market status per individual. Without considering matching and its related issues, we are considering all individuals within 2004Q1-2014Q4 and aged between 15 and 64 years who have at least three consecutive interviews available. Within this subset, we count all records which show a reported labour market pattern of either  $U - I - U$  or  $I - U - I$  in any three consecutive interviews available, irrespective the other labour market states in the remaining interviews<sup>17</sup>. In the Austrian LFS we find around six thousand individuals with such a vague differentiation between unemployment and inactivity. Again, compared to the around 330,000 people interviewed over the whole sample period, a negligible group. We may not correct for these transitions as we do not expect this group would bias our transition estimates along this dimension. However, omitting these cases can be used as robustness check (see Elsby et al. 2015).

Beside measurement errors at the U-I margin, Köhne-Finster & Lingnau (2008) report for the German LFS that many marginally employed (“geringfügig beschäftigt”) are often reported as inactive; either due to incorrect proxy interviews (e.g. a parent answers wrongly for their child), or respondents misunderstand the wording of relevant questions. As a result, this may also lead to spurious transitions along the employment and inactivity (E-I) margin, but without a clear prediction if this leads to an under- or over-reporting of worker flows between E and I. The literature, such as Petrongolo & Pissarides (2001) or Garibaldi & Wasmer (2005) deals with this flow in the manner of a „discrete-time“ issue. There, I-E flows are characterized by the misreporting of the intermediate unemployment (job search) state; typically thought to occur when individuals feature low job-search efforts<sup>18</sup>. However, Köhne-Finster & Lingnau (2008) has shown that misreporting along this dimension is not an issue of undetected search effort, but of misunderstanding the questionnaire and wrong answers in proxy interviews<sup>19</sup>. More specific to the Austrian LFS, the inactivity and employment margin ( $E - I$ ) is also characterized by large seasonal flows, as many seasonal workers drop out from employment and the most of them are then classified as inactive rather than unemployed<sup>20</sup>. Although, seasonal transitions may also be prone

<sup>17</sup> For this purpose, we used the wide (individual-specific) format of the raw dataset including the adjustments described above. Labour market patterns then are constructed, e.g. for an individual with five interviews  $E - (U - I - U) - U$ .

<sup>18</sup> Note that these individuals can exhibit a non-zero job-arrival rate

<sup>19</sup> See A.1 in the Appendix for some descriptives for the Austrian LFS.

<sup>20</sup> See section 5 for a graphical representation. In the Austrian LFS, the national feature of large seasonal *unemployment* fluctuations (national definition) turns into a large seasonal *inactivity* feature. This is related to the fact that the LFS applies ILO- rather than national definitions of labour market states. Hence, individuals are, although registered as unemployed nationally, classified as inactive if they are

for misreporting (seasonal jobs may be undeclared employment contracts), many seasonal jobs can be differentiated more precisely from inactivity than other (non-seasonal) jobs. This is, because the most seasonal jobs are either identified through personal interviews, or are typically not based on marginal employment contracts.

Although we can not provide a thorough analysis of the measurement of the  $E - I$  margin, the sources for misreporting errors stated in Köhne-Finster & Lingnau (2008) are also plausible for the Austrian LFS. Proxy interview rates are comparable to the German case and especially concentrated at the young; both facts that makes it plausible that inactivity and marginal employment is hardly differentiated. The  $E - I$  dimension may therefore suffer from misreporting, whereas it is not clear if we would detect too much or too less transitions. Unfortunately, we have to leave the reader without a solution to this issue. But, we conclude with the conviction that misreporting along this margin concerns only a minority of people in the data, as large seasonal worker flows are a special feature of the Austrian labour market which is by no means a result of misreporting.

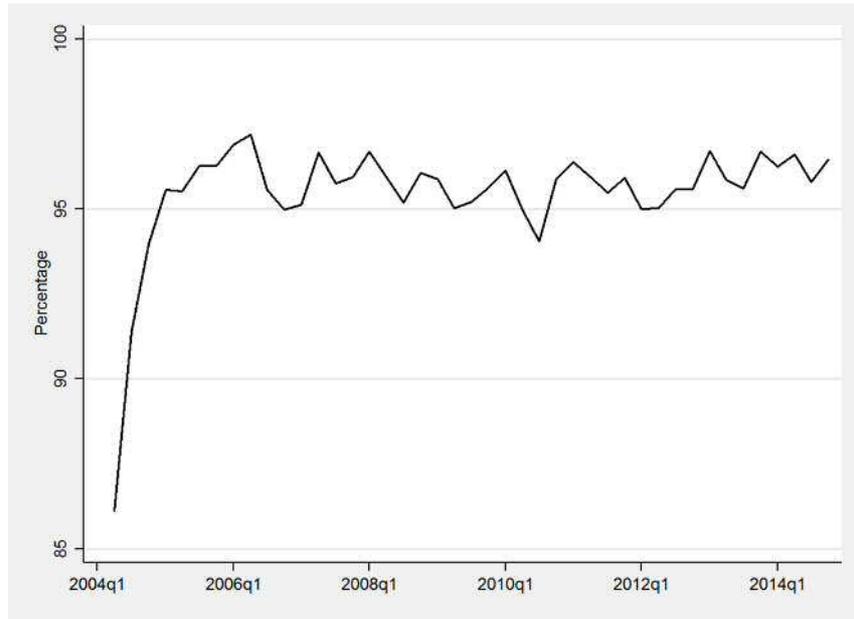
## 2.4 Non-response and matching probabilities

So far, we have obtained a crude two-quarter matched dataset of the Austrian LFS cross-sections which measures - at least for our purpose - the most important variables without significant biases. However, due to the rotational sampling scheme on the one hand, and individuals dropping from the sample on the other, we can only gather a subsample of the overall sample matching cross-sections. As a consequence, we need to have a close look on the consequences of this sample selection on the representativeness of the matched dataset. We do this (i) by looking at the origins of potential differences in individual matching probabilities through the lens of a discrete choice framework, (ii) by empirically testing if sample drop-outs cause a severe bias in our transition rate estimates  $\lambda_t^{ij}$ , and (iii) by restoring representativeness if the bias-test forces us to deal with an improper sample-selection.

As with the most labour force surveys, the Austrian LFS is also an address based survey. This means that addresses are sampled and every person eligible for the interview is surveyed at that address. Any person moving away will not be followed up by subsequent interviews, while new members locating themselves at that address will be included. A non-match between two consecutive interviews at time  $t - 1$  and  $t$  can therefore have several reasons: (i) migration, (ii) death, (iii) refusal, or (iv) other reasons such as the

---

not actively searching for a job or are not available for work. This is likely to be the case as many seasonal workers typically have a high recall-probability or a fixed agreement to become re-employed in the next season and are therefore not looking for a job.



**Figure 2:** Matching rate of individuals across two consecutive quarters (working-age population 15-64 years old, excluding the in-rotating group)

address does not exist any more<sup>21</sup>.

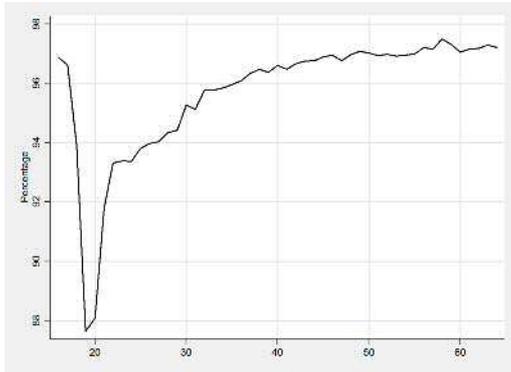
Figure 2 displays the matching rate over the period 2004Q2 to 2014Q4. On average we are able to match around 95.5 percent of individuals of a cross-section with the observations from the previous quarter. Interestingly, the year 2004 shows very low matching rates (90.5 percent) while the rate remains almost constant for all other quarters. The low rate in 2004 is due to the implementation phase of the survey and reflects difficulties to encounter each respondent from the previous period<sup>22</sup>.

Plotting the matching rate by several important variables reveals first informative insights of the matching determinants. The matching rate over the age profile in Figure 3a clearly indicates the high mobility of people around age twenty, while residential mobility typically decreases with age. As a result, the matching rate drops for youngsters around their home-leaving age. In Figure 3b, men typically show a slightly lower matching rate than women; potentially stemming from the fact that men are slightly more likely to refuse interviews in situations where no proxy-respondent is at hand. On average, to be an immigrant reduces the probability to be matched in comparison to natives. Figure 3c

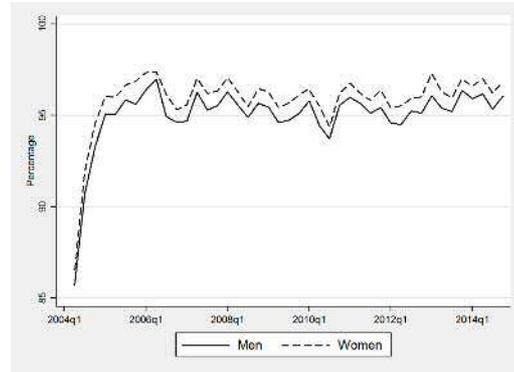
<sup>21</sup> Additionally, all individuals in the in-rotating group of quarter  $t$  can not be matched between quarters. However, as in-rotating groups always comprise random subsamples of each cross-section it is excluded in this subsection. We concentrate on systematic drop-outs/drop-ins only.

<sup>22</sup> We have further estimated all statistics in this article by excluding the year 2004. The results are available upon request. In general, it can be said, the fit of the models shown in Table 3 are more accurate, but remain qualitatively unchanged. The results of bias-testing and re-weighting also remain qualitatively unchanged

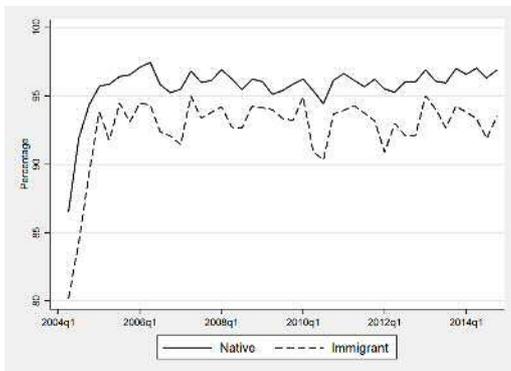
**Figure 3:** Matching rate by several variables



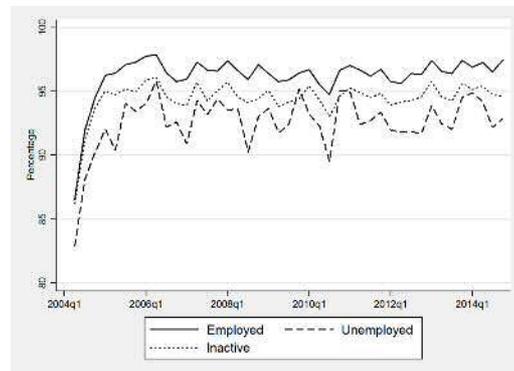
**Figure 3a:** Age



**Figure 3b:** Sex



**Figure 3c:** Citizenship



**Figure 3d:** Labour market status

shows that the average matching rate of immigrants is about three percentage points lower. It seems that it is much more difficult for the interviewers to encounter immigrants compared to natives<sup>23</sup>. Lastly in Figure 3d, the matching rate by labour market state in  $t$  clearly shows the higher drop-out rates for unemployed and also reflects higher mobility rates to take up jobs elsewhere. To sum up, the descriptives reveal some key-differences between individual matching probabilities which seems primarily driven by migration and the individual willingness to respond to the interview.

To look at the individual probability to become matched, we employ a logit model to analyse the effects of various covariates. Non-response and therefore non-matching in the Austrian LFS is a Markov process which means that the probability to respond in the interview at time  $t + 1$  only depends on the response in the current interview but not on

<sup>23</sup> Statistik Austria employs several interviewers with multiple language skills. It is therefore expected that the lower willingness to respond among immigrants is not due to a higher language barrier.

the interview in  $t - 1$  (see Table 8 in the Appendix). As a result, the probability to be matched between  $t$  and  $t + 1$  does not depend on lagged response behaviour. The binary dependent variable is one if an individual could have been matched between  $t$  and  $t + 1$  and zero otherwise<sup>24</sup>.

We estimate several specifications of the model reported in Table 3. The regressions include several individual (e.g. sex, age, etc.) and household (e.g. size of the household, number of children, etc.) characteristics. We further use a dummy for the date of the current interview, a dummy for the wave and one for the quarter within the year to control for effects stemming from the interview process. The same analysis but with a different dependent variable (matching between  $t - 1 - t$  instead of between  $t - t + 1$ ) is presented in the last two columns.

The results suggest that the highest matching probabilities are exhibited by individuals interviewed in the fourth quarter within the year and currently in wave four, being a native female 60-64 years old and employed, with tertiary education and unmarried within a single household together with children. This seems plausible, as the most variables indicate a low locational mobility and therefore a high chance to become re-interviewed. Furthermore, the gender difference in the matching probabilities remain after controlling for several other characteristics; women have - *ceteris paribus* - a 0.38 percentage points higher probabilities than men<sup>25</sup>. However, the gap is small. A much stronger impact stems from the immigrant status, the age of the respondent, the household structure, and also from being in the unemployment state. All this characteristics signal either a high propensity to move away from the sampled address, or a cultural influence on the response behaviour. These results are robust to altering the binary dependent variable from a forward to a backward-looking nature.

---

<sup>24</sup> We excluded those individual which do not have any interview in  $t$  and  $t + 1$  as we would only have information on time-invariant characteristics if interviews from other waves are available.

<sup>25</sup> Marginal effects, or differences in matching probabilities are estimated as differences in average predicted probabilities. As a further robustness check, we also estimated marginal effects as the difference in predicted probabilities of an average individual. All results remain unchanged, e.g. the gender-gap shrinks to 0.35 percentage points. The results are available upon request.

**Table 3:** Matching probabilities: Logit estimates of marginal effects

	Matching between $t$ and $t + 1$		Matching between $t - 1$ and $t$	
	Full sample	$\geq 2005Q1$	Full sample	$\geq 2005Q1$
<b>Wave</b> (Base: wave 1/wave 2)				
Wave 2	0.831***	0.767***	-	-
Wave 3	1.064***	0.954***	-0.781***	-0.639***
Wave 4	1.395***	1.265***	-0.622***	-0.535***
Wave 5	-	-	-0.474***	-0.480***
<b>Quarter</b> (Base Q1)				
Q2	0.144	0.307***	-1.403***	-0.547***
Q3	0.167**	0.150*	-1.093***	-0.764***
Q4	0.722***	0.754***	0.117*	0.264***
<b>Women</b>	0.384***	0.324***	0.514***	0.480***
<b>Age category</b> (Base: 15-19)				
20-24	-1.589***	-1.573***	-1.107***	-1.123***
25-29	-0.757***	-0.852***	1.222***	1.120***
30-34	0.386**	0.326*	2.439***	2.287***
35-39	1.279***	1.278***	2.781***	2.678***
40-44	1.896***	1.913***	3.241***	3.026***
45-49	2.468***	2.442***	3.773***	3.525***
50-54	2.847***	2.853***	3.978***	3.719***
55-59	3.319***	3.304***	4.393***	4.094***
60-64	3.489***	3.359***	4.528***	4.226***
<b>Labor market status</b> (Base: employed)				
Unemployed	-2.242***	-2.351***	-1.969***	-1.870***
Inactive	-0.626***	-0.739***	-0.780***	-0.792***
Night-empl.	-0.320***	-0.373***	-0.588***	-0.505***
Weekend-empl.	0.002	0.001	-0.252***	-0.236***
<b>Immigrant</b>	-3.653***	-3.684***	-2.433***	-2.325***
<b>Highest education</b> (Base: primary education)				
Secondary	0.191**	0.170**	0.044	0.003
Tertiary	0.297***	0.307***	0.367***	0.329***
<b>HH-type</b> (Base: Single)				
Couple with children	1.145***	1.117***	0.427***	0.462***
Single with children	3.391***	3.363***	2.490***	2.302***
<b>HH-size</b>	-0.230***	-0.224***	-0.049*	-0.071**
<b>Married</b>	-0.584***	-0.587***	-0.879***	-0.795***
Baseline probability	94.99	94.99	95.94	96.53
Pseudo-R2	0.05	0.05	0.05	0.06
Observations	1072455	933081	1007691	955758
$\hat{L}_m$	-203552	-176934	-161850	-135415
$\hat{L}_0$	-213370	-185425	-171221	-144154

Remarks: \*\*\* $p < 0.001$ , \*\* $p < 0.01$ , \* $p < 0.05$ . Binary logit estimation of matching probability. Marginal effects (i.e. averaged predicted probabilities) are reported. Matching is defined as two interviews between consecutive quarters; Not reported are the dummies for the date of first interview, region of residence, status within HH, no. of children and interactions of age and immigrant status.

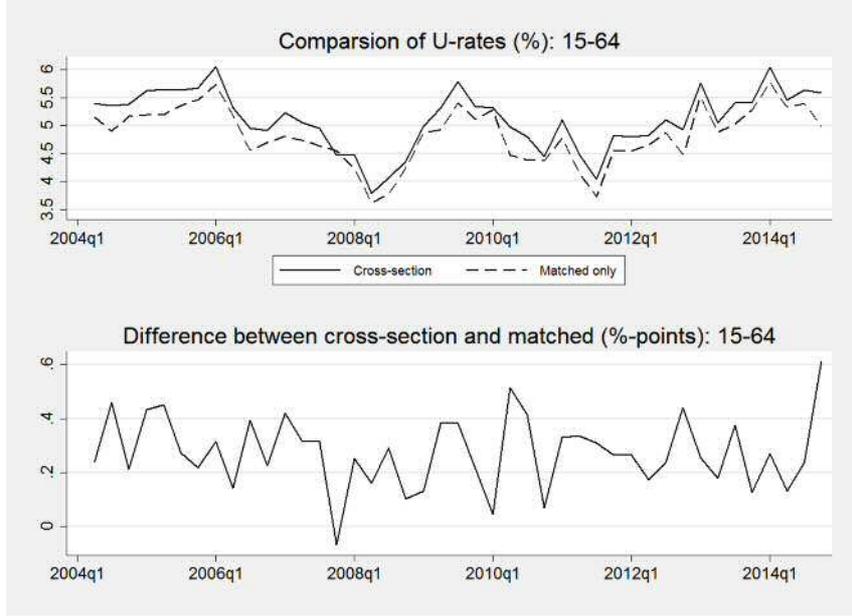


Figure 4: Comparison of the unemployment rate for the matched and unmatched dataset

### 3 Non-response bias

So far, we have shown that the probability to be matched between two consecutive quarters significantly differs between individuals. In this sections we aim to draw a conclusion on the consequences of these differences on the estimation of worker flows and transition rates if we exclude all individuals who can not be matched. We do this, by raising the question, if the selection process into the matched dataset is also related to estimates of labour market dynamics. If this is true, then the selection process also causes biased worker flow or transition estimates.

As shown in Figure 4, simply comparing the unemployment rate between the full and the matched dataset reveals markable differences. The idea of the statistical test for the presence of a bias is similar to this eye-balling exercise - we compare the transition rates as defined in (1) between the matched and the unmatched sample. Non-response introduces a bias if

$$\lambda_{(1)|x}^{ij} = E_x [Y_t = j | Y_{t-1} = i, D_k = 1] \neq E_x [Y_t = i | Y_{t-1} = j, D_k = 2] = \lambda_{(2)|x}^{ij} \quad (2)$$

where  $D_k$  is an indicator variable reflecting how the respondent  $k$  has participated in the survey. More specifically, we differentiate two states, where  $D_k = 1$  indicates that the respective respondent participated in the survey for two consecutive quarters (matched records) and  $D_k = 2$  identifies respondents who cannot be matched with the available

previous interview (response in  $t - 1$ , but non-response in period  $t$ ). The subscript  $x$  and the expectations operator  $E_x$  indicate that the equation is conditional on specific controls, or stratification variables  $X = x$  and that the expectation is taken conditional on the stratification  $X$ . We will deal with the choice of these stratification variables later.

More general, consider the unobserved population transition  $\lambda_x^{ij}$  which we would be able to estimate if we had all information on the non-matched records. This population transition rate can be decomposed into a part of the matched and a part of the unmatched data:

$$\begin{aligned}\lambda_x^{ij} &= \sum_{m=1}^2 E_x [Y_t = j | Y_{t-1} = i, D_k = m] P_x [D_k = m | Y_{t-1} = i] = \\ &= \sum_{m=1}^2 \lambda_{(m)|x}^{ij} P_x [D_k = m | Y_{t-1} = i].\end{aligned}\quad (3)$$

Equation 3 can be rearranged for the expected number of matched records being in labour market state  $j$  in  $t$  conditional on having been in state  $i$  in  $t - 1$  and assuming that  $P_x [D_k = 1 | Y_{t-1} = i] > 0$ , meaning that at least some records are matched.

$$\begin{aligned}E_x [Y_t = j | Y_{t-1} = i, D_k = 1] &= \lambda_x^{ij} \frac{1}{P_x [D_k = 1 | Y_{t-1} = i]} - \\ &- E_x [Y_t = j | Y_{t-1} = i, D_k = 2] \frac{P_x [D_k = 2 | Y_{t-1} = i]}{P_x [D_k = 1 | Y_{t-1} = i]}.\end{aligned}\quad (4)$$

As mentioned above,  $\lambda_x^{ij}$  and the expected value of the unmatched are unknown. However, we can find empirical counterparts of equation 4 and estimate the unknowns. The expectation of the matched (left-hand side) is the average number of individuals in state B conditional on having been in A before. The probability quantities are the shares of the respective group  $m = 1, 2$  in the total number of individuals of the population (or sample) having been in state  $i$ . Hence, the empirical counterpart can be written as

$$\begin{aligned}\lambda_{(1)|x}^{ij} &= \lambda_x^{ij} \frac{1}{P_{t-1}^{(1)}/A_{t-1}} - \lambda_{(2)|x}^{AB} \frac{A_{t-1}^{(2)}/A_{t-1}}{A_{t-1}^{(1)|i}/P_{t-1}^i} = \\ &= \beta_1 \frac{1}{P_{t-1}^{(1)|i}/P_{t-1}^i} + \beta_2 \frac{P_{t-1}^{(2)|i}}{P_{t-1}^{(1)|i}}.\end{aligned}\quad (5)$$

Testing if the observed transition rates of the matched are equal to the true population transition rates and to those of the unmatched, is equal to testing for equality of  $\beta_1 = -\beta_2$ . Hence, the testing procedure is a test for absence of non-response bias in our matched data set.

Concerning the stratification variables  $X$ , we choose some of the variables which have been identified as important determinants of non-response in the previous section. Unfortunately, our choice is limited as we would run into a small-cell problem when increasing the number of variables. Although this leads to a trade-off between the number of stratification variables for bias-reduction and the number of observation within a stratum, the choice is guided by a minimum amount of stratification variables to reduce the non-response bias to a negligible magnitude. Therefore, we choose sex, and three age groups (15-29, 30-49, 50-64 years) for natives and two age groups (15-44, 45-64) for foreign, in addition to the labour market status in  $t$  and  $t - 1$  (employment, unemployment, and inactivity). We estimate equation 5 by OLS for each stratum using the variation in the date of the  $t - 1$ -interview. The estimation procedure is performed by using unweighted and weighted counts, with weights being the official cross-section weights provided by Statistics Austria<sup>26</sup>. The p-values of the equality tests are shown in Table 4. The first column shows the results without stratification. For  $E \rightarrow E$ ,  $U \rightarrow E$ , and  $U \rightarrow U$  transition rates we see a significant non-response bias. Considering our stratification, we can conclude that we are not forced to reject the null-hypothesis of equal coefficients on average (last column). Hence, given our choice of the stratification variables, our transition rate estimates are (on average) not biased by non-response anymore.

## 4 Reweighting

In the previous section we have shown that in the matched dataset of the Austrian LFS some non-response bias is present. However, if transition rates are calculated sex-, age-group-, and citizenship-specific, we can conclude that the non-response bias is small enough to be statistically insignificant. To use the matched dataset of the Austrian LFS for several other types of research, it is necessary to provide a general tool to correct for the bias in the dataset. For this reason, we calculate specific *longitudinal weights* which could be used in combination to the provided survey weights to correct for systematic non-response.

In this section, we present a short discussion of different procedures how to obtain longitudinal weights in general and argue in favour of the intuitive iterative raking procedure, which is heavily used by statistical offices and thoroughly documented in Frazis et al.

---

<sup>26</sup> The results for weighted counts are not presented here but are similar and available upon request.

**Table 4:** Test for panel non-response bias: significance levels

$\lambda_x^{AB}$	Without stratification	Native men			Native women			Immigrant		Average
		15-29	30-49	50-64	15-29	30-49	50-64	15-44	45-64	
E → E	0.047	0.068	0.804	0.711	0.052	0.100	0.025	0.382	0.021	0.270
E → U	0.355	0.779	0.307	0.655	0.642	0.089	0.646	0.309	0.419	0.481
E → I	0.551	0.531	0.067	0.819	0.101	0.116	0.097	0.912	0.337	0.373
U → E	0.095	0.134	0.316	0.914	0.270	0.644	0.827	0.380	0.051	0.442
U → U	0.064	0.767	0.210	0.404	0.048	0.907	0.908	0.118	0.114	0.434
U → I	0.448	0.173	0.579	0.320	0.096	0.616	0.040	0.381	0.546	0.344
I → E	0.496	0.126	0.246	0.062	0.902	0.730	0.519	0.039	0.084	0.338
I → U	0.478	0.396	0.459	0.218	0.799	0.828	0.667	0.581	0.059	0.501
I → I	0.458	0.135	0.348	0.846	0.659	0.489	0.506	0.001	0.966	0.494

*Remarks: The table shows two-sided p-values of a F-test for equality of the two OLS-coefficients in equation (5). Results are obtained with unweighted counts (alternatively, similar results are obtained with weighted counts by using current period cross-section weights). Sample selection: 15-64yrs excluding people in military service.*

(2005) for the CPS. In general, the procedure aims to fit a system of worker flows between  $t - 1$  and  $t$  (based on the matched dataset) to respective labour market stocks in  $t - 1$  and  $t$  (based on cross-section datasets) by finding appropriate adjustment factors (longitudinal weights). After briefly sketching the used setup, we lay out the implemented procedure in detail. The raking procedure is then employed for each stratum and results in final longitudinal weights which serves as a bias-correction device for any future analyses. Furthermore, this section (and the respective code-files) contains general information how any matched version of the Austrian LFS can be adjusted to be appropriate for empirical analyses.

Many statistical offices provide special weights additionally to the cross-section weights or completely separate longitudinal weights (Kalton & Flores-Cervantes 2003). Both procedures incorporate the same idea of using weights to make the smaller longitudinal data set representative for certain population distributions and thereby to correct for systematic non-response in certain dimensions (see Kalton & Flores-Cervantes 2003, Deville et al. 1993, for a survey). Typically, statistical offices can directly obtain such weights in their population-weighting procedure. However, we can only obtain longitudinal weights *additional* to the available cross-section weights. While the cross-section weights correct for unit-non-response only in the respective cross-section, our additional weights aim to correct also for non-response in a dataset comprised of two matched surveys.

Given specific marginal population distributions, obtaining weights is equal to finding factors which make the marginal sample distributions conform to the population equivalent.

		Current quarter: $t$					$\Sigma_r$	
		E	U	I	OA	OB		OC
Previous quarter: $t - 1$	E	<i>EE</i>	<i>EU</i>	<i>EI</i>	<i>EOA</i>	<i>EOB</i>	<i>EOC</i>	$E_{t-1}$
	U	<i>UE</i>	<i>UU</i>	<i>UI</i>	<i>UOA</i>	<i>UOB</i>	<i>UOC</i>	$U_{t-1}$
	I	<i>IE</i>	<i>IU</i>	<i>II</i>	<i>IOA</i>	<i>IOB</i>	<i>IOC</i>	$I_{t-1}$
	AA	<i>AAE</i>	<i>AAU</i>	<i>AAI</i>	0	0	0	$AA_{t-1}$
	AB	<i>ABE</i>	<i>ABU</i>	<i>ABI</i>	0	0	0	$AB_{t-1}$
	AC	<i>ACE</i>	<i>ACU</i>	<i>ACI</i>	0	0	0	$AC_{t-1}$
$\Sigma_c$		$E_t$	$U_t$	$I_t$	$OA_t$	$OB_t$	$OC_t$	

**Table 5:** Structure of gross flows in the Austrian LFS

According to Kalton & Flores-Cervantes (2003), such (re-)weighting procedures can in general be divided into raking procedures and logistic methods. Logistic methods estimate a model for the probability to respond dependent on several explaining variables. The weights are then equal to the inverse of the predicted probabilities of the model for each respondent (see Lepkowski et al. 1989). Logistic methods have the advantage that they can deal with a large set of potential variables determining whether a sampled individual responds or not. In contrast, raking is an iterative adjustment procedure. It adjusts several sample figures iteratively by fitting the corresponding (marginal) sample distributions in accordance to the population distributions.

We decided to use the raking method as it is more intuitive, simple to implement, and easily adapted for other purposes. However, raking requires an a priori choice of marginal distributions, or stratification variables. As in the previous section, we have to deal with the trade-off between bias reduction (more stratification) and the number of stratum-specific observations (less stratification).

In our specific case, the raking procedure corrects the unadjusted worker flows to fit the current and the previous quarter cross-section labour market stocks. Again, we restrict our sample to individuals aged between 15 and 64 years. We also account for other in- and outflows into the longitudinal sample, not only because of attrition and re-entry. As a starting point we orientate on the adjustment procedure of Frazis et al. (2005) for the CPS. Their procedure can be schematically plotted as in Table 5. The matched data set

comprises only the shaded 3x3 matrix of worker flows, while the row and column totals represent the cross-section stocks of the respective quarter. The  $OA$ ,  $OB$ , and  $OC$  are outflows from the matched data set between  $t - 1$  and  $t$  due to out-rotation, attrition, and transiting out of our sample due to the 65<sup>th</sup> birthday.  $AA$ ,  $AB$ , and  $AC$  indicate inflows into the matched sample due to in-rotation, re-entering after attrition, and transiting in due to the 15<sup>th</sup> birthday<sup>27</sup>.

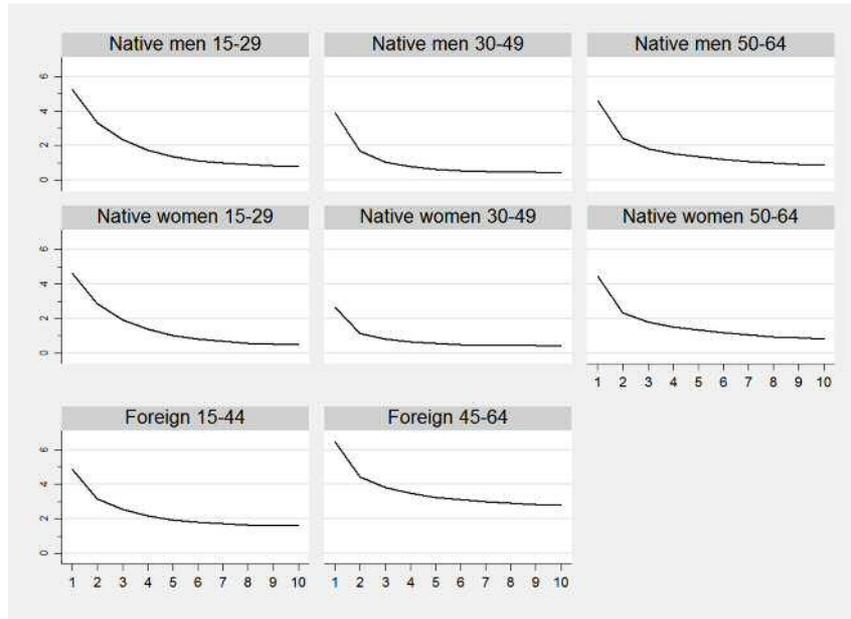
The raking procedure should ensure that both stock dimensions, those in  $t - 1$  and  $t$ , are met between the matched and the cross-sectional data set. For this purpose we have to assume that the out- and inflows  $O$ . and  $A$ . occur randomly with respect to the gross worker flows. The assumption is reasonable for the in- and out-rotating flows, as they comprise a random sub-sample in each quarter. For non-response we can draw on the results of the previous section, meaning that the choice of reasonable stratification variables reduces non-response to a random selection process. Additionally, we assume that this results hold in the same manner for attrition and re-entry flows. As a result, we choose the sex, the three age groups, the citizenship and the labour market states in both quarters as stratification variables with respect to the non-response flows. Considering the flows that are related to age boundaries, we conclude that they are negligible in comparison to the stock variables<sup>28</sup>. Our raking exercise is then done within each sex-age cell separately. To start the procedure we calculated the gross worker flows (shaded area), weighted with the current-quarter survey weights. Additionally, we obtain the cross-section stocks, weighted with the weights for the respective quarter. After that step our raking procedure has the following iterative steps:

1. Row-adjustment step: Proportional fitting is applied to the gross worker flows (shaded area) row-wise, so that row-totals meet the cross-section stocks for the previous quarter.
2. Update step: Column-margins are updated by summing column-wise.
3. Column-adjustment step: After 1, proportional fitting is applied such that the new column-totals meet the cross-section stocks of the current quarter.
4. Update step: Update row- and column-margins
5. Iterate till the algorithm converges.

---

<sup>27</sup> Military service is a separate labour market state but is not accounted for as such; it is treated as a form of attrition.

<sup>28</sup> All flows are far below 0.1 percent of the respective stock variable. Although covered in our analysis, we conclude that they have not the potential to introduce any biases. Furthermore, we rule out between-stratum flows, which of course are possible between age-groups. However, these flows are negligible.



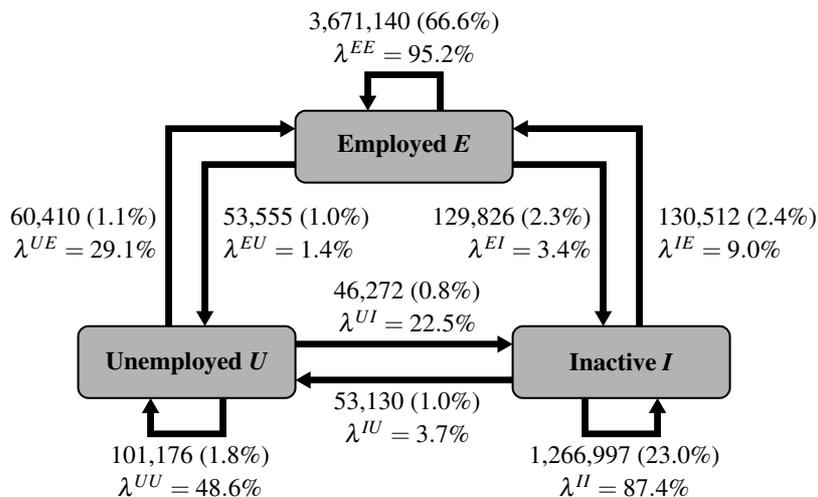
**Figure 5:** Relative row-margin error over iterations by strata [%]  
 Relative error is the absolute maximum of all labour market states. For each strata, averages over quarters are plotted. Working-age population (aged 15-64 years) only.

Convergence is reached if the absolute maximum of the column-wise relative errors are less than 0.00002 percent and the row-wise errors less than 10 percent<sup>29</sup>. Concerning convergence of the iterative procedure, there are no general criteria to ensure that the algorithm-errors converge towards zero. However, looking at the margin-errors over iterations, column-margin errors are virtually zero for all strata and quarters, because it is the last step in the iterative adjustment procedure. This last step adjusts the column-margins to perfectly correspond with the respective population figures. Therefore, convergence only has to be proofed along the row-dimension. The respective error-rates are depicted in Figure 5. The figure reports error rates in each iteration calculated as the absolute maximum across labour market states and as averages over quarters. The analysis of the tendency of the error-rates reveals that convergence is generally met for all combinations, although it is reached better for natives than for foreigners<sup>30</sup>.

Finally, longitudinal weights for each of the nine worker flows and each quarter-strata cell can be obtained by dividing the final (adjusted) worker flow estimates by the initial

<sup>29</sup> Relative errors mean that the difference between the raked and the non-raked stocks are expressed as percentage of the non-raked figures. The errors are calculated in each iteration and for each strata-quarter cell. The maximum is then taken from the absolute values over all strata-quarter combinations. Hence, convergence means that all margin-errors for each strata-quarter combination are below the thresholds.

<sup>30</sup> An general extension of the used algorithm would be to allow different stopping criteria for each quarter-strata combination. This could improve the fit for all series.



**Figure 6:** Average quarterly worker flows and transition rates (2004-2014); Relative share of working-age population (aged 15-64 years) in parentheses.

unadjusted worker flow estimates. As a result, the longitudinal weights are only quarter-, strata-, and worker-flow- specific series and turn only into individual weights when multiplied with the cross-section weights. These products are then longitudinal weights which can then be used for panel analysis correcting for sample selection biases due to non-response.

## 5 Worker flow estimates

After matching consecutive cross-sections and applying the longitudinal weights obtained through iterative raking, we are able to present unbiased worker flows and transition rates for the Austrian labour market. Quarterly worker flows  $N_t^{ij}$  are calculated as the (weighted) number of people transiting between labour market states  $i$  and  $j$ . Transition rates  $\lambda_t^{ij}$  are then calculated according to equation (1) and present a measure for the probability to leave a certain labour market state.

Given our results, the Austrian labour market is characterized by large worker flows. Figure 6 gives a first glance at the labour reallocation process using figures averaged over the period 2004Q2-2014Q4 and restricted to the population aged 15-64 years. On average, a number of 473,705 persons or 8.6% of the working age population changed their labour market status from one quarter to the other. Within the active labour force, 53,555 employed persons lose or quit their jobs while 60,410 unemployed persons find new jobs. Accordingly, worker flows between employment and unemployment typically reduce the pool of unemployment. We estimate an average job loss rate  $\lambda^{EU}$  of 1.4% and a job finding

**Table 6:** Comparison of labour turnover rates

Country	Period	H	S	LT	Country	Period	H	S	LT
AT	2005-2010	21.2	20.0	41.2	US	2005-2010	-	-	42.0
	2011-2014	18.5	18.1	36.6		2011-2014	-	-	38.0
ES		27.3	29.2	56.5	EE		15.8	13.4	29.3
SL		25.0	26.4	51.4	HU		14.6	12.2	26.8
FI	2011-2014	25.2	25.2	50.5	RO		12.9	13.2	26.2
PT		24.3	26.0	50.3	NL		12.6	13.0	25.7
DK		23.5	23.4	46.9	UK		11.9	10.7	22.6
LV		22.7	21.7	44.3	BG	2011-2014	10.9	11.5	22.4
<b>AT</b>	<b>2011-2014</b>	<b>19.5</b>	<b>18.9</b>	<b>38.4</b>	LT		11.4	10.0	21.4
CH		19.8	18.4	38.2	HR		9.6	11.3	20.9
IT		18.4	18.7	37.1	PL		10.0	9.4	19.4
SE	2011-2014	18.4	17.0	35.5	CZ		9.1	8.6	17.7
FR		16.1	15.5	31.6	GR		5.8	11.2	17.1
IE		15.3	14.9	30.3	SK		6.7	6.2	12.8
CY		13.9	15.9	29.7					

*Remarks: H ... Hiring rate, S ... Separation rate, LT ... labour turnover rate. Expressed as percentage of employment.*

*Source: First AT-entry own calculations based on LFS AT, European countries (including AT) from Eurostat (without Germany, Belgium, and Luxembourg), US from BLS based on JOLTS.*

rate  $\lambda^{UE}$  of 29.1%. Despite these sizeable flows within the labour force, the Austrian labour market also displays large movements from and into inactivity. Since participation increased substantially during the observed time period we detect that worker flows from inactivity into employment and unemployment exceeded the opposite flows leaving the labour force. Concerning the pool of unemployed again, inflows from inactivity typically increases unemployment.

Our estimates can also be compared to recently published figures by Eurostat. For the period 2010Q1 to 2014Q4, these figures offer worker flow estimates for European countries using the national Labour Force Surveys and a procedure similar to ours. This data source provides the most consistent measures for comparison with our newly acquired estimates. Only for illustrative purposes, Table 6 sums up labour turnover ( $LT_t = H_t + S_t$ ) which is the sum of total hirings ( $H_t$ ) and separations ( $S_t$ ) related to employment. We define hirings as worker flows into employment ( $N_t^{UE} + N_t^{IE}$ ) and separations as worker flows out of employment ( $N_t^{EU} + N_t^{EI}$ ). Within the pool of 25 European countries<sup>31</sup> the size of the Austrian labour turnover ranks at 7<sup>th</sup> position and is also of similar size compared to U.S. labour market<sup>32</sup>. Our estimates (first row) are comparable in size to

<sup>31</sup> Due to data limitations we are not able to present data for Germany, Belgium and Luxembourg.

<sup>32</sup> Estimates for the U.S. are based on the Job Openings and Labor Turnover Survey (JOLTS) provided by the Bureau of Labor Statistics (BLS).

those published by Eurostat. Although, the Eurostat data is also based on individual matching of subsequent cross-sections, it does so without establishing the same standards of longitudinal consistency presented in this study for the Austrian case.

After focusing on the total size of the Austrian reallocation system, we briefly consider the dynamics of flows and rates over time and their cyclical characteristics. Figure 7a and Figure 7b display the evolution of gross worker flows and transition rates, with and without seasonal adjustment. At the participation margin, we find strong seasonal patterns within the employment-inactivity-margin suggesting that these transitions are affected by seasonal fluctuations in the production process of certain industries (see the unadjusted series). These movements lead to strong seasonality of the overall participation rate which is not the case for the unemployment rate since flows from and into unemployment only feature less pronounced seasonal patterns (refer to footnote 20). This phenomenon of the Austrian labour market is exceptional compared to many other countries and should be further investigated to analyse if this is driven by compositional and/or institutional effects. In order to remove these seasonal fluctuations, we apply the X-12-ARIMA-method with multiplicative seasonal factors. We find statistical evidence for stable or moving seasonality at different magnitude for all worker flows except for the transitions from unemployment into inactivity<sup>33</sup>. For consistency reasons we suggest to seasonally adjust all worker flows whenever the time series are used to analyse the cyclical dynamics of employment, unemployment and inactivity.

Regarding the cyclicity of employment and unemployment it is useful to analyse the underlying reallocation patterns using the estimated transition rates. Starting with the job loss rate  $\lambda^{EU}$  and the job-finding rate  $\lambda^{UE}$ , we find a negative relation between both over the period 2004-2014. After job loss reached its minimum and job finding peaked after a pronounced expansionary phase in the middle of 2008, both rates responded to the starting economic turmoil in the aftermath of the Financial Crisis pushing up unemployment rates. Based on the seasonally adjusted data the job loss rate increased by almost 50 percent between the second quarter 2008 and the second quarter of 2009 while the job finding rate dropped by around 20 percent during the same period. Subsequently, job finding picked up pace while job loss declined during the short period of recovery between 2010 and 2011. Since then, the Austrian unemployment rate and the job loss probability show an increasing trend. In the same period, the job finding component dropped again as a reaction to the low growth rates of the Austrian economy. This strong decrease in job finding clearly led to increasing upward pressure of Austrian unemployment.

Beside job loss and job-finding transitions, Austrian (un-)employment is particularly shaped by transitions from and into inactivity. At the beginning of our observation period

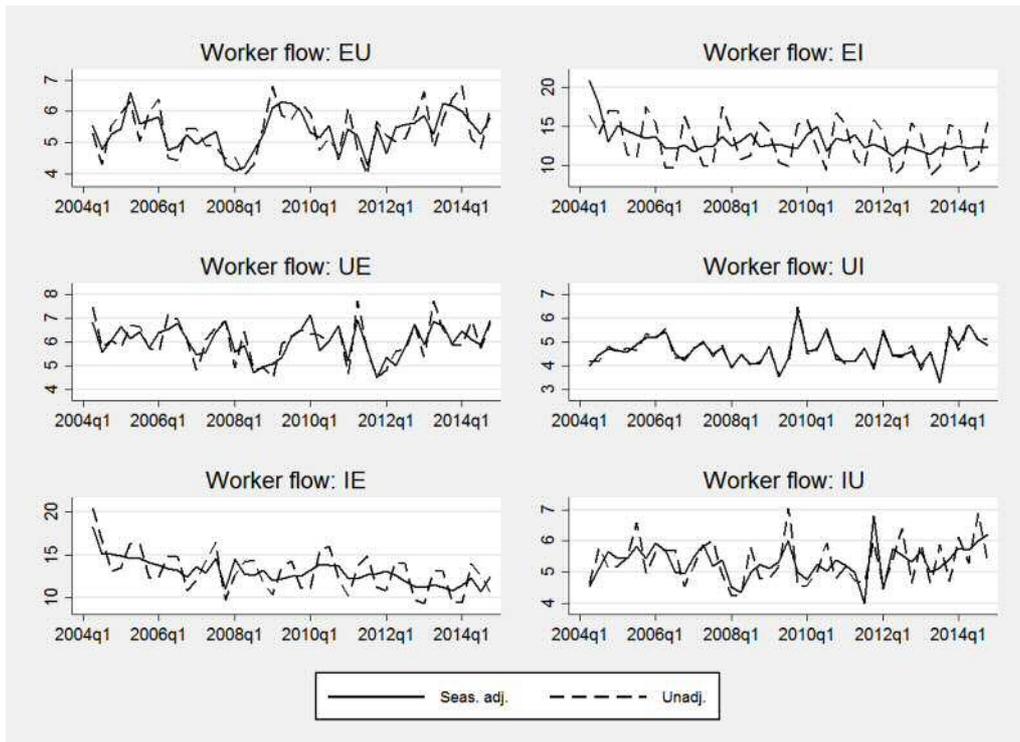
---

<sup>33</sup> A full documentation of test results for all worker flows and all strata is available upon request.

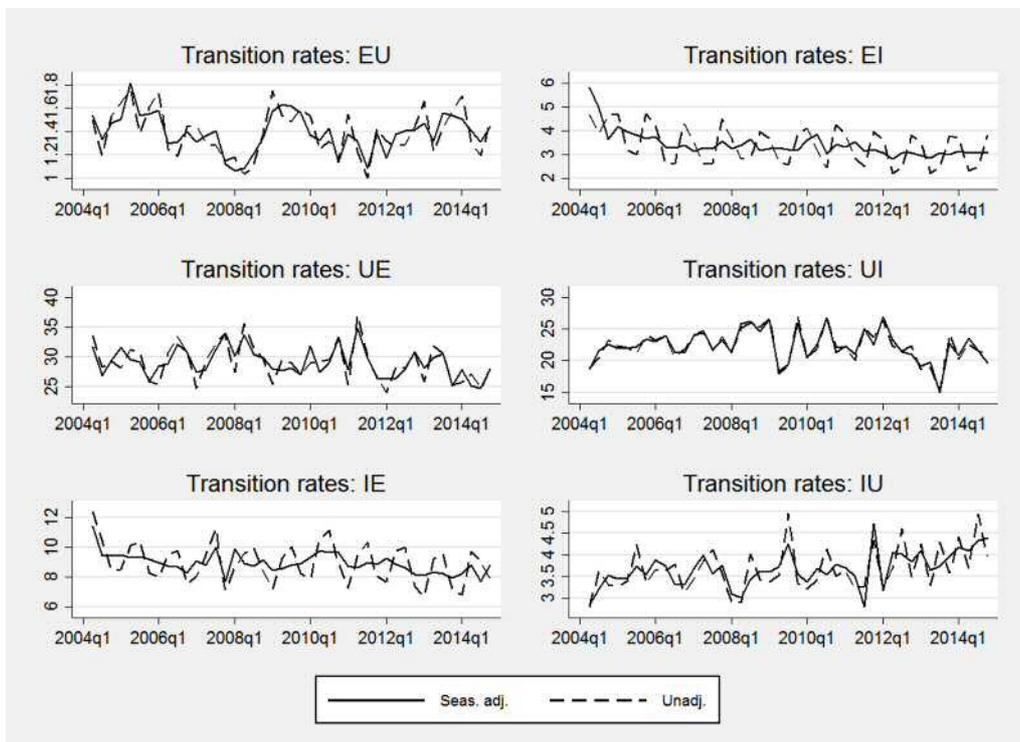
we notice large movements especially at the employment-inactivity margin. Both, flows and rates have strongly declined from 2004 to 2007. This drop was largely triggered by an exceptional increase of participation of workers over the age of fifty<sup>34</sup>. In this context the  $I - U$  margin is apparently more related to the economic cycle compared to  $I - E$  margin. During to the first phase of the crisis,  $I - U$  transitions rose while  $U - I$  transitions fell - these effects further increased the unemployment rate at that time. Interestingly, both transition rates show enhanced variability after the crisis and signal much stronger labour force attachment after 2011. This leads to the impression that out-of-labour-force-fluctuations have become a more important component for total unemployment dynamics in recent years.

---

<sup>34</sup> Note that large scale changes in labour market policy and reforms of the pension system took place during this period.



**Figure 7a:** Worker flows  $N_t^{ij}$ . In 10,000 people, weighted by cross-section and longitudinal weights. Seasonally adjustment with multiplicative X12-ARIMA.



**Figure 7b:** Transition rates  $\lambda_t^{ij}$ . In percent of initial stock  $P_{t-1}^i$ , weighted by cross-section and longitudinal weights. Seasonally adjustment with multiplicative X12-ARIMA.

## 6 Conclusion

This article constructs an important pre-requisite for worker flow analysis in Austria by using the rotational panel structure of the Austrian Labour Force Survey (LFS/Mikrozensus) from 2004Q1-2014Q4. It enables researchers to construct an appropriate longitudinal two-quarter matched dataset for the analysis of labour market dynamics. For this purpose, we provide a coherent procedure which incorporates several important issues when constructing such a dataset.

Matching individuals between two consecutive quarters often involves misreporting in several key variables - in particular in labour market variables - which in turn can lead to biased worker flow estimates (Abowd & Zellner 1985). Although we are not able to perform an in-depth analysis of measurement issues, we conclude that concerns arise more likely at the inactivity-employment margin rather than at the unemployment-inactivity side. We argue that this might be caused by the potential lack of knowledge of proxy-interviewees (mostly parents) over marginal employment contracts of other household members (e.g. children) if they mainly perform other activities than employment (e.g. study). We further conclude that extensive analysis would be necessary in this respect, but that the variables are accurately measured for our purposes.

Constructing the matched dataset constitutes a selection process as some individuals could not be matched due to non-response. In general, we lose around four to five percent of each quarterly sample. By applying a statistical test, we show that this leads to a sample-selection bias in transition rate estimates. Within a discrete-choice analysis on the determinants of individual matching probabilities, we carve out important variables of individual non-response behaviour and show that a stratification by sex, citizenship and age leads to unbiased worker flows and transition rates.

Based on this stratification and the conclusion that the sample-selection bias is removed when conditioning on this set of stratification variables, we apply an intuitive iterative raking procedure to ensure consistency between cross-section stocks in the current and previous quarter to their respective flow equivalents. The result of this procedure are longitudinal weights which serve as a bias-correction device for any future analysis.

Using these weights we estimate time series on worker flows and transition rates for the period 2004Q2-2014Q4. Our results suggest that Austria is characterized by relatively large worker flows between labour market states compared to other European countries. Within the labour force, worker flows between employment and unemployment are negatively related and display high sensitivity to economic shocks. Furthermore, we find sizeable movements within the participation margin. In this regard, we highlight the

substantial increase of participation, the strong seasonal patterns of worker flows between employment and inactivity and enhanced variability of transitions between unemployment and inactivity in the aftermath of the Financial Crisis.

## References

- Abowd, J. & Zellner, A. (1985), 'Estimating gross labor force flows', *Journal of Business and Economic Statistics* **3**, 254–283.
- Aumayr, C. (2010), 'Inter- and intraindustrial Job-to-Job Flows', *Review of Economic Analysis* **2**, 86–109.
- Bachmann, R., Bechara, P., Kramer, A. & Rzepka, S. (2015), 'Labour market dynamics and worker heterogeneity during the great recession - evidence from europe', *IZA Journal of European Labor Studies* **4**(19).
- Deville, J.-C., Sarndal, C.-E. & Sautory, O. (1993), 'Generalized raking procedures in survey sampling', *Journal of the American Statistical Association* **88**(423), 1013–1020.
- Elsby, M., Hobijn, B. & Sahin, A. (2010), 'The labor market in the great recession', *Brookings Papers on Economic Activity* **1**, 1–48.
- Elsby, M., Hobijn, B. & Sahin, A. (2015), 'On the importance of the participation margin for labor market fluctuations', *Journal of Monetary Economics* **72**, 64–82.
- Elsby, M. W., Hobijn, B. & Sahin, A. (2013), 'Unemployment dynamics in the OECD', *Review of Economics and Statistics* **95**(2), 530–548.
- Elsby, M. W., Michaels, R. & Solon, G. (2009), 'The ins and outs of cyclical unemployment', *American Economic Journal: Macroeconomics* **1**(1), 84–110.
- Eurostat (2015), *Labour market transitions - LFS longitudinal data Metadata*.  
**URL:** [http://ec.europa.eu/eurostat/cache/metadata/en/lfsi\\_long\\_esms.htm](http://ec.europa.eu/eurostat/cache/metadata/en/lfsi_long_esms.htm)
- Flek, V. & Mysíková, M. (2012), Unemployment dynamics in central europe: A labor flow approach, Technical report, Charles University Prague, Faculty of Social Sciences, Institute of Economic Studies.
- Frazis, H., Robison, E., Evans, T. & Duff, M. (2005), 'Estimating gross flows consistent with stocks in the cps', *Monthly Labor Review* **9**, 3–9.
- Frijters, P. & van der Klaauw, B. (2006), 'Job search with non-participation', *The Economic Journal* **116**, 45–83.
- Fujita, S. & Ramey, G. (2009), 'The cyclicalities of separation and job finding rates', *International Economic Review* **50**, 415–430.
- Garibaldi, P. & Wasmer, E. (2005), 'Equilibrium search unemployment, endogenous participation, and labor market flows', *Journal of the European Economic Association* **3**(4), 851–882.
- Gomes, P. (2010), 'Labour market flows: Facts from the united kingdom', *IZA DP No. 5327*.
- Gumprecht, D. & Oismüller, A. (2013), 'Non-response im Mikrozensus', *Statistische Nachrichten* **11/2013**, 1046–1061.
- Hausman, J. & Wise, D. (1979), 'Attrition bias in experimental and panel data: The gary income maintenance experiment', *Econometrica* **47**, 455–473.
- Hofer, H., Pichelmann, K. & Schuh, A.-U. (2001), 'Price and quantity adjustments in the austrian labour market', *Applied Economics* **33**, 581–592.

- ILO (1982), Resolution concerning statistics of the economically active population, employment, unemployment and underemployment, *in* ‘adopted by the Thirteenth International Conference of Labour Statisticians’.
- Jimenez-Martin, S. & Peracchi, F. (2002), ‘Sample attrition and labor force dynamics: Evidence from the spanish labor force survey’, *Spanish Economic Review* **4**, 79–102.
- Kalton, G. & Flores-Cervantes, I. (2003), ‘Weighting methods’, *Journal of Official Statistics* **19**(2), 81–97.
- Köhne-Finster, S. & Lingnau, A. (2008), ‘Untersuchung der Datenqualität erwerbsstatistischer Angaben im Mikrozensus Ergebnisse des Projekts: Nachbefragung im Mikrozensus/LFS’, *Statistisches Bundesamt - Wirtschaft und Statistik* **12/2008**, 1067–1088.
- Lepkowski, J., Kalton, G. & Kasprzyk, D. (1989), Weighting adjustments for partial nonresponse in the 1984 sipp panel, *in* ‘Proceedings of the American Statistical Association, Section on Survey Research Methods’.
- Meraner, A., Gumprecht, A. & Kowariak, A. (2015), ‘Die neue hochrechnung des mikrozensus: Methodenbeschreibung’, *Statistik Austria Working Paper* .
- Mitterndorfer, B., Schrittwieser, K. & Stadler, B. (2007), ‘Analyse von panelausfi $\frac{1}{2}$ llen im mikrozensus’, *Statistische Nachrichten* **7**, 616–625.
- Nordmeier, D. (2012), ‘Worker flows in germany: Inspecting the time aggregation bias’, *IAB Discussion Paper 12/2012* .
- Peracchi, F. & Welch, F. (1995), ‘How representative are matched cross-sections? evidence from the current population survey’, *Journal of Econometrics* **68**, 153–179.
- Petrongolo, B. & Pissarides, C. A. (2001), ‘Looking into the black box: A survey of the matching function’, *Journal of Economic Literature* **39**, 390–431.
- Petrongolo, B. & Pissarides, C. A. (2008), ‘The ins and outs of European unemployment’, *The American Economic Review* **98**(2), 256–262.
- Poterba, J. & Summers, L. (1986), ‘Reporting errors and labor market dynamics’, *Econometrica* **54**, 1319–1338.
- Shimer, R. (2005), ‘The cyclical behavior of equilibrium unemployment and vacancies’, *American Economic Review* **95**(1), 25–49.
- Shimer, R. (2012), ‘Reassessing the ins and outs of unemployment’, *Review of Economic Dynamics* **15**, 127–148.
- Statistik Austria (2015), *Mikrozensus Erlä $\frac{1}{2}$ uterungen 2015*.
- Stiglbauer, A. (2003), Job and Worker Flows in Austria 1978-98, PhD thesis, University of Linz.
- Stiglbauer, A., Stahl, F., Winter-Ebmer, R. & Zweimüller, J. (2003), ‘Job creation and job destruction in a regulated labor market: The case of austria’, *Empirica* **30**(2), 127–148.
- van den Berg, G., Lindeboom, M. & Ridder, G. (1994), ‘Attrition in longitudinal panel data, and the empirical analysis of dynamic labor market behaviour’, *Journal of Applied Econometrics* **9**, 421–435.

## A Appendix

### A.1 Measurement errors due to interview characteristics

For the Austrian LFS, the proxy interview rate is comparable to similar surveys in other countries. E.g. Köhne-Finster & Lingnau (2008) report an average proxy rate of one quarter which is close the reported 27 percent in Austria (see overall average in Figure 7). This average is clearly driven by high rates during the implementation phase of the new survey design in 2004. For this period, misreporting along the inactivity and employment margin is much more likely. According to Table 7, the largest share of proxy interviews for the young (15-29 years) are given by their parents, while proxy interviews for individuals over 30 are typically conducted with the individual's partner. All in all, the high proxy rate for the young give some indication that marginal employment among the young might be inadequately reported, while the stable and well-below-average rate after the implementation phase does not indicate any distortions due to the type of the interview.

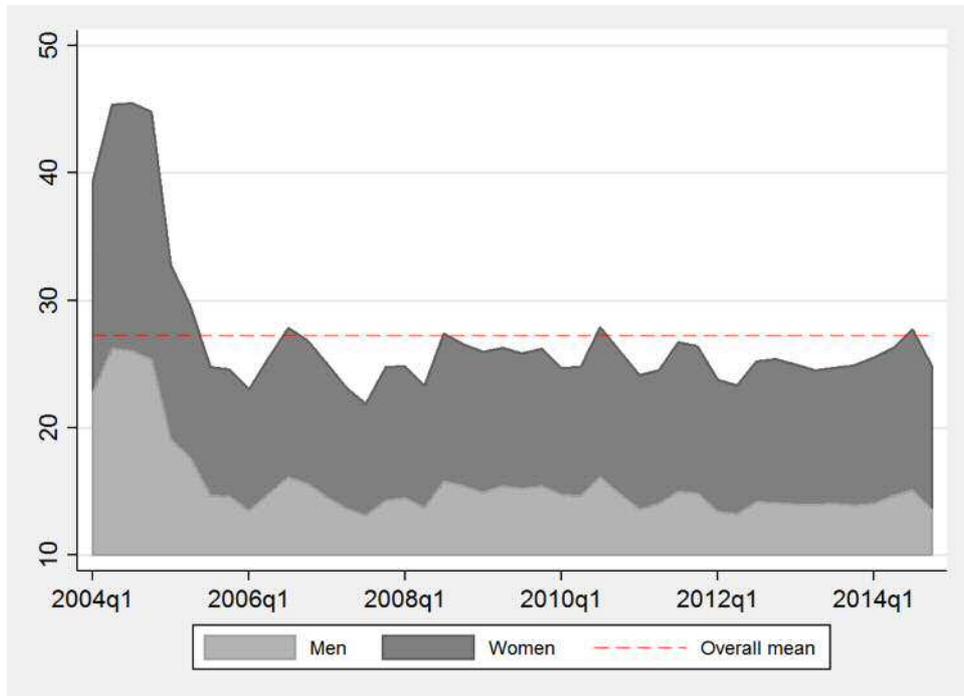


Figure 7: Proxy interview rate over time

Proxy type	Age			<b>Total</b>
	15-29	30-49	50-64	
Personal	53.72	78.93	79.90	<b>72.72</b>
Proxy: Partner	4.35	15.80	15.41	<b>12.74</b>
Proxy: Mother	25.25	1.75	0.27	<b>7.38</b>
Proxy: Father	12.28	0.84	0.07	<b>3.56</b>
Proxy: Other	4.40	2.68	4.35	<b>3.61</b>

*Remarks: Restricted to working-age (15-64 years) only. Total observations: 1,376,303*

**Table 7:** Percentage type of proxy interview by age

## A.2 Non-response probabilities over waves

Table 8 shows the overall probabilities that an individual does not respond in the next interview in  $t+1$ . The unconditional probabilities do not vary to a large extent and display only minor increases till the end of the whole survey period. The conditional figures give the respective quantities dependent on whether an individual participated in the current  $t$  interview or not. Furthermore it also conditions on participation in the previous  $t-1$  and the current  $t$  interview. The conditional figures show that the participation lagged two-periods ( $t-1$ ) does not substantially alter non-response probability in  $t+1$ . However, participation in the current interview does. This means that (non-)response behaviour of individuals can be well described by a Markov process of order one without controlling for more than one lag. This is important for modelling matching probabilities within a discrete choice framework, as the probability is the conditional probability to respond in both, the current and the next interview. In this situation it is not necessary to further control for response behaviour in the lagged interview.

		Wave of interview				
		1	2	3	4	5
Unconditional		14.9	16.8	17.5	18.0	18.4
Conditional on						
$t$	0	67.8	69.5	71.6	73.4	-
$t$	1	7.7	6.8	6.6	6.2	-
$t - 1, t$	00	-	64.2	68.2	71.2	-
$t - 1, t$	10	-	77.4	78.5	78.2	-
$t - 1, t$	01	-	9.5	11.6	11.4	-
$t - 1, t$	11	-	6.7	6.2	5.9	-

*Remarks: Sample 2004Q1-2014Q4, age 15-64, excluding people in military service (first available observation); 0...non-response, 1...response.*

**Table 8:** Probabilities of non-response in the interview in wave/quarter  $t + 1$