I Z A

IZA DP No. 3908

# Inconsistencies in Reported Employment Characteristics among Employed Stayers

Francesca Bassi
Alessandra Padoan
Ugo Trivellato

December 2008

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

# Inconsistencies in Reported Employment Characteristics among Employed Stayers

## Francesca Bassi
*University of Padova*

## Alessandra Padoan
*Regione Veneto*

## Ugo Trivellato
*Universtiy of Padova,*
*CESifo and IZA*

# ABSTRACT

# Inconsistencies in Reported Employment Characteristics among Employed Stayers

The paper deals with measurement error, and its potentially distorting role, in information on industry and professional status collected by labour force surveys. The focus of our analyses is on inconsistent information on these employment characteristics resulting from yearly transition matrices for workers who were continuously employed over the year and who did not change job. As a case-study we use yearly panel data for the period from April 1993 to April 2003 collected by the Italian Quarterly Labour Force Survey. The analysis goes through four steps: (i) descriptive indicators of (dis)agreement; (ii) testing whether the consistency of repeated information significantly increases when the number of categories is collapsed; (iii) examination of the pattern of inconsistencies among response categories by means of Goodman's quasi-independence model; (iv) comparisons of alternative classifications. Results document sizable measurement error, which is only moderately reduced by more aggregated classifications. They suggest that even cross-section estimates of employment by industry and/or professional status are affected by non-random measurement error.

Corresponding author:

Ugo Trivellato
Faculty of Statistical Science
University of Padova
Via Cesare Battisti 241
35121 Padova
Italy
E-mail: trivell@stat.unipd.it

## 1.    Introduction and summary[*]

In recent years, labour markets in industrialised countries have shown quite a high degree of mobility. Extensive literature on the micro-dynamics of the labour market focuses on job-to-job flows (Steven and Haltiwanger, 1999; Fallick and Fleischman, 2004; Shimer, 2005, among many others). Only a few papers explore the kind of job changes that workers experienced. The complexity of job mobility also demands analysing these changes while changing employer (Neal, 1999). The literature on job matching suggests that a significant number of workers who switch job also change employment characteristics, mainly industry and occupation (Miller, 1984; McCall, 1990).

Some studies show that job characteristics, particularly industry and occupation, collected in surveys are affected by measurement error. The effect of these errors is to exaggerate the occurrence of changes in such characteristics, at least when information is obtained at two points in time with independent interviews (Bound, Brown and Mathiowetz, 2001). Sala and Lynn (2006) compare estimates of changes in industry and occupation obtained in two survey waves 17 months apart but with different interview techniques: traditional independent interviewing *vs.* dependent interviewing. They show that dependent interviewing results in lower levels of observed changes and that this shrinkage represents a reduction in measurement error, since the effect is particularly pronounced among respondents who do not change job between waves. Other studies demonstrate that, in general, industry is reported more accurately than occupation and that, not surprisingly, the agreement rate between employees' and employers' reports, classified according to a single-digit coding scheme, is higher than that resulting when reports are categorised according to the more detailed three-digit classification (Mellow and Sider, 1983; Mathiowetz, 1992).

In this paper we deal with measurement error, and its potentially distorting role, in information on industry and professional status[1]. As a case-study we consider two-wave panels one year apart collected by the Italian Quarterly Labour Force Survey (QLFS) in the period from April 1993 to April 2003.

In Italy, information on job characteristics can be obtained from various sources. As a prominent example, WHIP (Work Histories Italian Panel), a panel built from a sample of micro-data from the administrative archives of the Italian social security agency (INPS),

[1] As we will show in Section 2.1, the classification used by the Italian Quarterly Labour Force Survey questionnaire to collect information about occupation is a collapsed mixture of two classifications: by "occupation", as defined by the International Standard Classification of Occupation, and by "status of employment", as defined by the International Classification of Status of Employment (see ILO, 1993 and 2008, respectively). Following Eurostat (2000), from now on we will call it "professional status".

provides longitudinal information on work histories of employees in the private non-agricultural sector. Using data for the period 1985-1996 and a three-digit ATECO81 classification[2], Leombruni and Quaranta (2002) show that 53% of job changes coincide with industry changes. Mobility among occupations, measured according to a 4-category classification[3], is lower: 7% of blue-collar workers and executives change occupation while changing job; the percentage rises to 17% for white-collar workers. Updated estimates from the same authors (Leombruni and Quaranta, 2005) document the persistence of these patterns.

Another fundamental source for the analysis of short-term dynamics and persistence in the Italian labour market is the QLFS. It has the distinctive advantage of referring to a sample of the resident non-institutional population. Thus, it collects information on job characteristics of (almost) all the employed. The survey is cross-sectional with a 2-2-2 rotating design, which yields two-wave panels one quarter and one year apart (see, e.g., Trivellato, 1997). The focus of our analyses is on inconsistent information on employment characteristics – industry and professional status – resulting from yearly transition matrices for workers who reported that they were continuously employed over the year and did not change job.

First, we compute and comment upon some usual indicators of disagreement. We find clear evidence that there is sizable measurement error in both industry and professional status, and that industry is reported more accurately than professional status. We then expand our analysis in three directions: (i) we test whether the consistency of repeated information, provided by employment stayers, significantly increases when the number of categories is collapsed; (ii) we explore the pattern of inconsistencies among response categories using Goodman's (1968) quasi-independence model; (iii) we compare the appropriateness of alternative classifications jointly by professional status and industry.

As regards the detail of variable classification for cross-section estimates (admittedly less demanding than estimates from two-wave panel data), Istat – the Italian statistical agency – provides the following indications. For professional status, a reliable classification reduces to a binary one: Employee and Self-employed. For industry, Istat asserts as dependable a classification in 12 categories. Based on the hierarchical Kappa coefficient, for each of the two variables, industry and professional status, we test if reducing the number of categories significantly increases the consistency of information reported in two interviews one year apart. Evidence from these analyses supports the first indication by Istat, but casts severe doubts on the second. Significant results in terms of measurement error reduction are obtained for a 6- or 5-category classification of industry.

We further explore the patterns of inconsistencies among categories of variables by testing several specifications of Goodman's quasi-independence model, which is almost always rejected.

---

[2] ATECO81 is the old Italian version of NACE (*Nomenclature statistique des Activités économiques dans la Communauté Européenne*).

[3] The 4-category classification for occupation consists of Executives, White-collars, Blue-collars and Apprentices.

Lastly, we consider and compare alternative 4-category classifications obtained by collapsing professional status and industry into a single variable. The standard classification labels respondents as Self-employed, Employee in agriculture, Employee in industrial sector, and Employee in services. As an alternative, another 4-category classification was recently used by Trivellato *et al.* (2005) in their study of worker turnover, with the four categories given by Self-employed, Employee in agriculture, Employee in industrial sector and private services, and Employee in Public Administration and social services. Interestingly enough, the latter classification turns out to be almost uniformly better than the former, standard one.

The paper proceeds as follows. Section 2 contains a brief description of the data and presents the methods and design of the analyses. Section 3 reports the main results. Section 4 concludes.

## 2. Data and methods

### 2.1. Data

As already mentioned, QLFS is a quarterly survey with a 2-2-2 rotating design[4]. It collects information about labour market participation on a sample of respondents from the resident non-institutional population. For all persons who declare themselves as employed or report that they worked at least one hour during the reference week, the questionnaire includes a series of questions on employment characteristics: working hours, professional status, industry, place of work, number of persons working at the local unit, type of contract, and date in which person started working for the current employer or in the current activity.

Professional status is identified by means of a closed-form question with 11 categories, 6 for employees (Manager, Executive, Clerk, Workman, Apprentice, Outworker) and 5 for self-employed (Entrepreneur, Professional, Own-account worker, Member of a producers' cooperative, Contributing family worker). Information on industry is collected with an open-ended question and coded by the interviewer according to the ATECO2002[5] classification. As already noted, Istat (2003) warns that these two variables may not be totally reliable, if used at their maximum degree of detail. It suggests using the binary classification, Employee/Self-employed, for professional status, and a 12-category classification, corresponding to the two-digit ATECO2002, for industry – the 12 categories are listed in Table 4, column 1.

---

[4] The description of the survey given here applies until 2003. In 2004, the survey was substantially redesigned. The main change regards the timing of interviews: while up to 2003 QLFS interviews took place only in one week – usually the second – of each quarter, with the new Continuous LFS interviews spread out over all weeks of the quarter. Other notable changes regard the largely new questionnaire and the mode of interview: from paper and pencil to CAPI-CATI (Istat, 2004).

[5] ATECO2002 is identical to NACE Rev. 1.1 at four-digit level.

In this paper we use 10 two-wave yearly panels, from April 1993 to April 2003. Among workers who were continuously employed during the year – yearly employment stayers - we consider only those who did not change job, for a total of 263,584 sample units, around 25,000 for each panel. We adopted a quite strict criterion for identifying these workers: they are those respondents who, in both one-year-apart interviews, were classified as employed and reported the same date when answering the question: "When did you start working with the current employer or in the current self-employment?"[6].

Following de Angelini and Giraldo (2003), we consider inconsistencies in job characteristics reported one year apart as due to measurement errors affecting these variables, and assume that no or negligible errors are made in reporting dates. We also assume that, among yearly employment stayers who did not change employer or current self-employment, genuine levels of change in industry and/or professional status were likely to be very low (Sala and Lynn, 2006). Thus, operationally for those workers all observed change in industry and/or professional status is attributed to measurement error (Mathiowetz and McGonagle, 2000).

As regards professional status, measurement error is likely to be due to the detailed classification offered to respondents; for industry, mainly to the nature of the open-ended question used to collect information. Minor changes in the wording used by respondents to describe their branch of activity or in recoding information by the interviewer might have led to a different industry classification, when in fact no change occurred.

*2.2. Methods and design of analyses*

Our study consists of analysing inconsistencies in industry and professional status reported in two independent interviews one year apart by workers who were (reasonably assumed to be) continuously employed and did not change job. The analysis develops along several lines.

First, the usual descriptive indicators to assess inconsistencies are computed and compared across the two variables and over time.

Transition matrices among job characteristics declared one year apart provide the basic information for quantifying inconsistencies. As an example, Tables 1 and 2 report the transition matrices by industry and professional status, respectively, of the April 1993-April 1994 panel. The frequencies on the main diagonal refer to consistent responses, while those outside the main diagonal point to inconsistencies.

A simple indicator of disagreement is the percentage of frequencies outside the main diagonal ($P$), or gross difference rate as defined by Hansen, Hurwitz and Pritzker (1964) in their seminal paper.

_____

[6] The criterion is conservative: in accordance with the aim of our work, it is meant to minimise the risk of including false positives. Thus, we decided to eliminate (i) records with missing data regarding dates, and (ii) records with dates which differed in the two interviews, although they were consistent with the fact that the worker was continuously employed during the year and did not change job.

**Table 1:** Transition matrix by industry, April 1993 to April 1994

| 1993 | 1994 | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Agri-colture | Min-ing | Manu-factur-ing | Con-struc-tion | Whole-sale | Ac-co-mod. | Trans-porta-tion | Finan-cial act.s | Profes-sional act.s | Public. admin. | Educa-tion | Other public act.s |
| Agricolture | 1,624 | 4 | 26 | 8 | 17 | 4 | 2 | 1 | 3 | 30 | 7 | 4 |
| Mining | 1 | 270 | 32 | 22 | 14 | 0 | 2 | 4 | 2 | 8 | 1 | 1 |
| Manufacturing | 16 | 28 | 5,315 | 87 | 179 | 15 | 41 | 11 | 43 | 29 | 37 | 36 |
| Construction | 12 | 12 | 90 | 1,721 | 28 | 0 | 16 | 6 | 27 | 17 | 12 | 21 |
| Wholesale | 24 | 6 | 171 | 55 | 3,648 | 40 | 31 | 6 | 22 | 19 | 28 | 34 |
| Accomodation | 2 | 1 | 6 | 5 | 26 | 705 | 10 | 0 | 0 | 6 | 15 | 14 |
| Transportation | 11 | 2 | 47 | 22 | 31 | 4 | 1,306 | 10 | 4 | 44 | 11 | 21 |
| Financial act.s | 2 | 5 | 17 | 8 | 21 | 2 | 10 | 777 | 12 | 12 | 11 | 10 |
| Profess. act.s | 2 | 2 | 39 | 39 | 32 | 3 | 9 | 23 | 814 | 22 | 22 | 65 |
| Public admin. | 23 | 11 | 37 | 22 | 25 | 10 | 48 | 13 | 12 | 2,098 | 164 | 27 |
| Educ.& helth | 7 | 2 | 25 | 10 | 28 | 13 | 10 | 10 | 11 | 108 | 3,188 | 38 |
| Other public & social act.s | 9 | 5 | 25 | 14 | 39 | 19 | 25 | 8 | 36 | 40 | 58 | 938 |

**Table 2:** Transition matrix by professional status, April 1993 to April 1994

| 1993 | 1994 | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Man-ager | Execu-tive | Clerk | Work-man | Appren-tice | Out-worker | Entre-preneur | Profes-sional | Own-account | Coop's member | Contr. family |
| Manager | 243 | 89 | 39 | 14 | 0 | 0 | 6 | 7 | 4 | 0 | 0 |
| Executive | 65 | 684 | 218 | 11 | 0 | 0 | 0 | 11 | 4 | 0 | 1 |
| Clerk | 55 | 284 | 6,790 | 465 | 0 | 2 | 13 | 31 | 44 | 8 | 12 |
| Workman | 15 | 16 | 557 | 7,882 | 25 | 11 | 8 | 6 | 138 | 22 | 35 |
| Apprentice | 0 | 0 | 11 | 79 | 98 | 1 | 0 | 1 | 2 | 0 | 1 |
| Outworker | 0 | 0 | 4 | 18 | 0 | 29 | 0 | 1 | 5 | 0 | 0 |
| Entrepreneur | 4 | 0 | 6 | 4 | 0 | 0 | 238 | 21 | 123 | 6 | 10 |
| Professional | 6 | 13 | 36 | 6 | 0 | 1 | 7 | 641 | 93 | 2 | 3 |
| Own-account w. | 4 | 4 | 40 | 118 | 2 | 6 | 102 | 89 | 4,353 | 74 | 102 |
| Coop's member | 0 | 0 | 6 | 14 | 0 | 0 | 8 | 3 | 54 | 115 | 7 |
| Contr. family w. | 0 | 3 | 36 | 36 | 5 | 1 | 12 | 7 | 117 | 14 | 767 |

Cohen's (1960) Kappa coefficient, $K = (p_o - p_e)/(1 - p_e)$, where $p_o$ is the observed proportion of agreement and $p_e$ is the proportion of agreement expected under independence of answers, is a well-established index of agreement. It is based on the comparison between observed counts on the main diagonal of the matrix with the corresponding expected cell counts under the model of independence. It ranges from –1 (total disagreement) to 1 (perfect agreement). Zero is obtained when agreement is totally due to chance[7][8].

Second, we want to ascertain if reducing the number of the categories in which the variables are coded significantly increases the consistency of information reported in two subsequent interviews. We test that by using the hierarchical Kappa coefficient (Cohen, 1968).

The hierarchical Kappa coefficient provides a framework for investigating if patterns of disagreement pertain primarily to interchanges among similar response categories, as opposed to substantively important misclassification. Partial credit is permitted for a certain type of disagreement, which implies assigning a set of weights to some specific matrix cells. For example, the weights can be chosen so that the associated Kappa measures indicate the increments in agreement which result by successively combining relevant categories of the response variable. In this case: $p_o = \sum_{i=1}^{I} \sum_{i'=1}^{I} w_{ii'} p_{ii'}$ and

$p_e = \sum_{i=1}^{I} \sum_{i'=1}^{I} w_{ii'} p_{i.} p_{.i'}$, where $w_{ii'}$ is the weight assigned to the $ii'$ cell, $p_{ii'}$ is the proportion

---

[7] Other indicators suggested by the literature, given a transition matrix with absolute frequencies, are (see, e.g., Hauser and Massagli, 1983): (i) The net difference rate $e_i = [(X_{.i} - X_{i.}) / X_{..}] \times 100$, where $X_{.i}$ is the $i$th marginal column sum, $X_{i.}$ is the $i$th marginal row sum, and $X_{..}$ is the total count. This expression is simply the difference between interviews in the marginal proportions in the same categories. It ranges from –100 to 100, 0 is obtained when marginal proportions for a category are exactly the same. Its limitation is that there may appear to be significant differences in marginal proportions in several categories as a result of a smaller number of net classification differences. (ii) The index of inconsistency: $I_i = \{(X_{.i} + X_{i.} - 2X_{ii})/[X_{.i}(X_{..} - X_{i.}) + X_{i.}(X.. - X_{.i})]/ X_{..}\} \times 100$, where $X_{ii}$ is the $i$th diagonal entry (for a critical review of this index, see Biemer, 2004). This is the ratio of observed discrepancies (off-diagonal counts) in a given category to those discrepancies expected under simple independence. It ranges from 0 – no inconsistencies – to 100 – complete randomness among answers. Its major defect is that the model of independence almost never fits repeated measurements. We calculated the net difference rate and the index of inconsistency on our transition matrices, and obtained results fully consistent with those revealed by indexes $P$ and $K$, shown in Table 3.

[8] Cohen's Kappa is especially appropriate in the medical sciences, where studies are often designed to assess the agreement between different raters or different diagnostic instruments. If the two readings are from two different raters, $K$ accounts for rater bias; if the two readings are from replicated measurements, an intraclass correlation coefficient is appropriate, since we may assume no bias (Barnhart and Williamson, 2002). In our case, we aim at measuring agreement between responses obtained on the same sample in interviews one year apart: the interviewer is not usually the same; in addition, the respondent might also change, since proxy respondents were allowed, and indeed frequent (see, e.g., Gandolfo and Gennari, 2000, who document that, in the period April 1998 to January 1999, the rate of proxy respondents in the QLFS was on average slightly over 40%). For these reasons, we prefer Cohen's Kappa to the intraclass correlation coefficient.

observed in cell *ii'* and $p_{i.}p_{.i'}$ is the proportion expected under the model of independence (Koch *et al.*, 1977).

The problem is formalised as a simple test of hypothesis. Let $\hat{K}_1$ and $\hat{K}_2$ be two hierarchical Kappa coefficients, estimated with two different sets of weights so that the second one implies a less disaggregated classification, the hypothesis test $H_o : \hat{K}_2 = \hat{K}_1$ *vs.* $H_1 : \hat{K}_2 > \hat{K}_1$ allows one to verify if aggregating categories improves significantly agreement (Landis and Koch, 1977). In our study, the weights $w_{ii'}$ are chosen so that they imply aggregation among categories identifying similar employment (industry or professional status) characteristics. The weights have value 1 for cells in which there is perfect agreement (on the main diagonal) and for cells outside the main diagonal linking similar categories – whose observed frequencies are considered as agreements, value 0 for all other cells.

Finally, the patterns of inconsistencies among response categories, at various levels of disaggregation, are explored by estimating log-linear models of quasi-independence.

Log-linear models can be usefully applied in order to detect inconsistencies in contingency tables (Hagenaars, 1990). In particular, the model of quasi-independence is used to evaluate if, leaving aside the main diagonal cells, the remaining cells show particular systematic patterns of association or whether there is independence on this (truncated) table. In the model of quasi-independence, the entries on the diagonal cells of a transition matrix are blocked, and the model of independence is specified for the off-diagonal cells (Goodman, 1968). The expression of the log-linear model of quasi-independence for an *IxI* table is the following:

$$\log F_{ij} = \mu + \mu_i + \mu_j + \mu_{ij},$$

where $F_{ij}$ is the expected frequency in the generic cell of the two-way contingency table; $\mu$ is the grand mean; $\mu_i$ are row effects with $1 \le i \le I$; $\mu_j$ are column effects with $1 \le j \le I$, and $\mu_{ij}$ are interaction effects for diagonal cells, $\mu_{ij} = 0$ if $i \ne j$.

In our case, assuming quasi-independence implies that errors in reporting industry or professional status are independent in two interviews one year apart. Rejecting the model implies that inconsistencies do not occur randomly: rather, there are systematic patterns of associations among response categories. After estimating the model, a close inspection of residual frequencies may give information on the sizes of associations.

## 3. Main results

### 3.1. Descriptive evidence

For the 10 two-wave yearly panels – from April 1993 to April 2003 – of yearly employment stayers who did not change job, as described in Section 2.1, Table 3 contains the values of the descriptive statistics of inconsistency: the percentage of frequencies

**Table 3:** Measures of inconsistencies with reference to industry and professional status

| Panels | Industry | | Professional status | |
|--------|----------|----------|---------------------|----------|
| | *P* * | *K*** | *P* * | *K*** |
| 93-94 | 11.8 | 0.8672 | 14.0 | 0.8132 |
| 94-95 | 10.6 | 0.8785 | 13.1 | 0.8255 |
| 95-96 | 10.9 | 0.8750 | 13.1 | 0.8265 |
| 96-97 | 9.5 | 0.8915 | 12.5 | 0.8349 |
| 97-98 | 9.7 | 0.8896 | 12.9 | 0.8297 |
| 98-99 | 10.6 | 0.8787 | 12.9 | 0.8301 |
| 99-00 | 10.9 | 0.8758 | 13.1 | 0.8276 |
| 00-01 | 10.9 | 0.8754 | 13.7 | 0.8195 |
| 01-02 | 10.3 | 0.8822 | 12.5 | 0.8346 |
| 02-03 | 9.7 | 0.8892 | 12.4 | 0.8355 |

\* *P* is the percentage of frequencies outside the main diagonal.
\*\* *K* is the Cohen's Kappa coefficient.

outside the main diagonal (*P*) and Cohen's Kappa (*K*), with reference to industry classified into 12 categories, as recommended by Istat, and to professional status classified into 11 categories, as in the questionnaire.

It is worth noting that industry is reported with fewer inconsistencies than professional status, according to both indices[9]. Another interesting piece of evidence is that there is no significant trend in the indices of inconsistencies: the effect of measurement error in the survey is fairly constant over the decade.

Looking at the indices calculated for the various categories of the two variables – not reported here for the sake of space[10] – it emerges that the most consistent categories of professional status are Clerk and Workman among employees, and Professional and Own-account worker among self-employed. In industry, Agriculture, Mining and raw material extraction, Professional and support service activities, and Other public, social and personal service activities are reported with the least inconsistencies.

In order to interpret the values of Cohen's Kappa, two scales are mainly used in empirical studies. The scale proposed by Fleiss (1981) defines as marginal agreement values of the coefficient which are lower than 0.40, good agreement values between 0.41 and 0.75, and excellent agreement values over 0.75. Landis and Koch (1977) propose to consider as slight agreement values lower than 0.20, fair agreement those between 0.21 and 0.40, moderate agreement values between 0.41 and 0.60, and substantial agreement in

---

[9] As already mentioned, this result can also be found in other studies. However, in our case study it is slightly puzzling, since professional status was reported by answering a closed-form question, whereas industry was asked by means of an open-ended question and answers were afterwards coded by the interviewer. Literature on measurement errors in surveys, specifically in reporting job characteristics, shows that inconsistencies over time are more likely when information is collected with open-ended questions (Mathiowetz and McGonagle, 2000).

[10] Available from the authors on request.

cases of values between 0.61 and 0.80. A value greater than 0.80 denotes almost perfect agreement.

As shown in Table 3, for both variables – industry and professional status – the values of Cohen's Kappa are quite high, resulting in almost perfect agreement according to both scales. All coefficients are also statistically different from zero. However, the peculiarity of our case should be taken into account. Indeed, in the absence of measurement error we expect no inconsistencies at all between information reported in two subsequent interviews. First evidence of non-negligible inconsistencies in our data is the percentage of frequencies outside the main diagonal of the matrices, around 10% for industry and 12-14% for professional status. In addition, under the hypothesis of no inconsistencies, one should expect a Kappa coefficiente equal to or very close to 1, which does not seem to be the case[11]. Overall, we interpret these results as indicative of sizable measurement errors.

We proceed therefore to verify if aggregating categories would improve agreement, i.e., significantly diminish the percentage of inconsistent information.

### 3.2. A strategy for testing a sequence of less disaggregated classifications

As already explained, the procedure based on hierarchical Kappa coefficients is appropriate for assessing the pattern of agreement among two or more classifications of some categorical response variables. A sequence of hierarchical Kappa coefficients refers to progressively less stringent, usually nested, definitions of agreement. The values of the coefficient obtained yield larger values for corresponding broader views of agreement. Since Kappa coefficients have an approximate multivariate normal distribution for large samples, it is possible to test the significance of successive differences by means of Wald statistics.

Hierarchical Kappas are formulated using sets of criterion weights. In our application, the first set of weights defines the agreement as the occurrence of the same response category in both interviews. The other sets of weights correspond to more aggregated classifications, which also consider as agreement the occurrence, in two consecutive interviews, of responses which are different but belong to similar categories. Similar categories are treated as equivalent, yielding to a less stringent definition of consistency. The aggregation process, described in Tables 4 and 6, was based both on judgment and on evidence on the distribution and size of inconsistencies across categories documented by the descriptive analyses.

For industry, the first set of weights is given by the 12-category classification recommended by Istat. The second set results in a 6-category classification, as shown in column 2 of Table 4. The third set of weights corresponds to a 5-category classification, derived from the previous one as shown in column 3 of Table 4. The last set of weights

---

[11] Note that it would be quite complicated to build a test-statistic to ascertain the hypothesis that the empirical Kappa coefficients were significantly different from 1, since it would imply testing a parameter value at the boundary of the parameter space.

**Table 4:** Scheme of categories aggregation process: industry

| 12 categories | 6 categories | 5 categories | 3 categories |
|---|---|---|---|
| Agriculture | Agriculture | Agriculture | Agriculture |
| Mining and raw material extraction | Manufacturing and mining | Manufacturing and mining | Industrial sector |
| Manufacturing | | | |
| Construction | Construction | Construction | |
| Wholesale and retail trade | Wholesale and retail trade | Wholesale and retail trade | Services |
| Accommodation and food services | Services | Other activities | |
| Transportation and communications | | | |
| Financial and real estate activities | | | |
| Professional and support service activities | | | |
| Public Administration, defence and compulsory social security | Public Administration | | |
| Education, health and other services | | | |
| Other public, social and personal service activities | | | |

**Table 5:** Hierarchical Kappa coefficients and Wald tests: industry

| Panels | Kappa coefficients | | | | Wald tests | | |
|---|---|---|---|---|---|---|---|
| | 12 categories | 6 categories | 5 categories | 3 categories | 6 *vs.* 12 | 5 *vs.* 6 | 3 *vs.* 5 |
| 93-94 | 0.8672 | 0.8833 | 0.8940 | 0.9020 | 217.75*** | 96.26*** | 26.59*** |
| 94-95 | 0.8785 | 0.8939 | 0.9037 | 0.9113 | 174.26*** | 71.25*** | 21.41** |
| 95-96 | 0.8750 | 0.8899 | 0.8989 | 0.9005 | 193.51*** | 71.79*** | 1.17 |
| 96-97 | 0.8915 | 0.9044 | 0.9104 | 0.9159 | 165.64*** | 38.77*** | 14.39 |
| 97-98 | 0.8896 | 0.8982 | 0.9044 | 0.9082 | 81.08*** | 34.85*** | 6.19 |
| 98-99 | 0.8787 | 0.8894 | 0.8964 | 0.9008 | 107.82*** | 40.62*** | 7.65 |
| 99-00 | 0.8758 | 0.8880 | 0.8931 | 0.8944 | 132.95*** | 21.90** | 0.65 |
| 00-01 | 0.8754 | 0.8865 | 0.8883 | 0.8903 | 109.25*** | 2.91 | 1.39 |
| 01-02 | 0.8822 | 0.8910 | 0.8926 | 0.8943 | 80.50*** | 2.47 | 1.08 |
| 02-03 | 0.8892 | 0.9008 | 0.9046 | 0.9044 | 131.86*** | 14.05 | 0.02 |

* Significant at α=0.1, ** significant at α=0.05, *** significant at α=0.01.

implies the usual 3-category classification: Agriculture, Industrial sector, and Services (column 4 of Table 4).

Table 5 contains the values of the Kappa coefficients calculated using the four sets of classifications by industry on the 10 panels, and the results of Wald tests performed on the differences between each hierarchical Kappa coefficient and the one corresponding to a more aggregated classification[12]. Switching from 12 to 6 categories significantly improves agreement among responses in all 10 panels; reducing further categories to 5 significantly improves agreement in 7 out of 10 panels; no significant increase is obtained when reducing answers to the usual 3-category classification. From these results, it appears that the two classifications that minimise inconsistencies in information on industry collected in the QLFS are those with 6 or 5 categories. This evidence challenges the Istat's recommendation to use the 12-category classification.

For professional status, the first set of weights consists of the 11-category classification used in the questionnaire. The second set results in a 6-category classification obtained by aggregation, as in column 2 of Table 6. The last set of weights corresponds to the binary classification recommended by Istat: Employee and Self-employed (column 3 of Table 6).

Table 7 is similar to Table 5, and contains the Kappa coefficient values calculated using the three sets of classifications by professional status on the 10 panels, as well as the results of the tests performed on the differences between hierarchical Kappa coefficients. Switching from 11 to 6 categories significantly improves agreement among responses in all

**Table 6:** Scheme of categories aggregation process: professional status

| 11 categories | 6 categories | 2 categories |
|---|---|---|
| Manager | White-collar | Employee |
| Executive | | |
| Clerk | | |
| Workman | Blue-collar | |
| Apprentice | | |
| Outworker | Outworker | |
| Entrepreneur | Self-employed | Self-employed |
| Professional | | |
| Own-account worker | | |
| Member of a producers' cooperative | Member of a producers' cooperative | |
| Contributing family worker | Contributing family worker | |

---

[12] Since Kappa coefficients have an approximated multivariate normal distribution for large samples, chi-square tests for linear hypotheses about them can be carried out with Wald statistics.

**Table 7:** Hierarchical Kappa coefficients and Wald tests: professional status

| Panels | Kappa coefficients | | | Wald tests | |
|---|---|---|---|---|---|
| | 11 categories | 6 categories | 2 categories | 6 *vs.* 11 | 2 *vs.* 6 |
| 93-94 | 0.8132 | 0.8709 | 0.9317 | 1,035.71*** | 621.99*** |
| 94-95 | 0.8255 | 0.8803 | 0.9371 | 816.87*** | 486.58*** |
| 95-96 | 0.8265 | 0.8804 | 0.9361 | 931.28*** | 560.05*** |
| 96-97 | 0.8349 | 0.8904 | 0.9402 | 965.87*** | 487.31*** |
| 97-98 | 0.8297 | 0.8850 | 0.9406 | 927.84*** | 552.70*** |
| 98-99 | 0.8301 | 0.8863 | 0.9398 | 922.88*** | 512.05*** |
| 99-00 | 0.8276 | 0.8816 | 0.9388 | 874.81*** | 565.95*** |
| 00-01 | 0.8195 | 0.8764 | 0.9317 | 893.42*** | 481.62*** |
| 01-02 | 0.8346 | 0.8911 | 0.9413 | 922.73*** | 466.68*** |
| 02-03 | 0.8355 | 0.8894 | 0.9432 | 893.83*** | 527.38*** |

*** Significant at α=0.01.

panels[13]; reducing categories to 2 further increases agreement. The Istat's recommendation to classify professional status with a dichotomous variable, Employee/Self-employed, is neatly confirmed.

### 3.3. Does the quasi-independence model hold?

In the case of industry, the log-linear model of quasi-independence was estimated so to reproduce the 12-, 6-, 5- and 3-category classifications. Table 8 lists the results of model fitting (Pearson $X^2$ and log-likelihood ratio $L^2$ statistics and associated *p*-values and BIC index) for the 10 panels. The hypothesis of quasi-independence is always rejected, for all panels and at all levels of aggregation, indicating that one-year-apart responses show non-random association even at the maximum level of aggregation (3 categories).

Although the model of quasi-independence does not fit, the BIC index gives some interesting results. It reaches its minimum value with the 6-category classification in 7 out of 10 panels, indicating that this is the level of aggregation which fits the data best, and with the 5-category classification in the remaining three panels. This evidence confirms the conclusions reached by applying the hierarchical Kappa procedure.

Estimated residuals of a quasi-independence model measure association exceeding that expected under random behaviour. The greater their value, the higher the difference between the observed association and that expected under the null hypothesis of randomness. Specifically, positive estimated residuals indicate that the model underestimates association between the two categories involved; negative estimated residuals indicate overestimation. Inspecting estimated residuals of the quasi-independence model implying 3 categories, i.e., testing quasi-independence among inconsistencies which

---

[13] In Section 2.1 we assumed that, among continuously employed workers who did not change job, genuine levels of change in professional status (and/or industry) were likely to be negligible. In any case, with the 6-category classification, true transitions among professional statuses for those workers are definitely implausible.

**Table 8:** Goodness-of-fit statistics for quasi-independence model: industry

| Panels | Number of categories | $X^2$* | $L^2$* | Degrees of freedom | BIC index |
|---|---|---|---|---|---|
| 93-94 | 12 | 1,310.95 | 1,180.35 | 109 | 74.87 |
|  | 6 | 651.81 | 596.52 | 89 | -306.12 |
|  | 5 | 350.64 | 308.96 | 65 | -350.27 |
|  | 3 | 202.72 | 190.25 | 47 | -286.42 |
| 94-95 | 12 | 1,117.23 | 993.54 | 109 | -93.73 |
|  | 6 | 557.06 | 514.57 | 89 | -373.20 |
|  | 5 | 234.53 | 219.63 | 65 | -428.75 |
|  | 3 | 111.99 | 108.30 | 47 | -360.52 |
| 95-96 | 12 | 1,073.58 | 965.40 | 109 | -138.19 |
|  | 6 | 520.16 | 501.48 | 89 | -399.61 |
|  | 5 | 253.15 | 240.34 | 65 | -417.77 |
|  | 3 | 145.81 | 145.19 | 47 | -330.67 |
| 96-97 | 12 | 1,038.38 | 896.49 | 109 | -206.29 |
|  | 6 | 411.06 | 393.80 | 89 | -506.63 |
|  | 5 | 220.96 | 212.57 | 65 | -445.05 |
|  | 3 | 117.53 | 119.77 | 47 | -355.74 |
| 97-98 | 12 | 776.94 | 692.80 | 109 | -407.74 |
|  | 6 | 359.91 | 365.29 | 89 | -533.31 |
|  | 5 | 240.31 | 229.31 | 65 | -426.97 |
|  | 3 | 111.58 | 111.95 | 47 | -362.59 |
| 98-99 | 12 | 990.77 | 887.54 | 109 | -210.85 |
|  | 6 | 511.34 | 489.06 | 89 | -407.79 |
|  | 5 | 277.61 | 273.25 | 65 | -381.76 |
|  | 3 | 138.15 | 142.21 | 47 | -331.40 |
| 99-00 | 12 | 1,087.13 | 997.56 | 109 | -102.56 |
|  | 6 | 542.33 | 524.01 | 89 | -374.25 |
|  | 5 | 284.80 | 273.11 | 65 | -382.92 |
|  | 3 | 183.87 | 174.50 | 47 | -299.86 |
| 00-01 | 12 | 976.55 | 861.55 | 109 | -234.43 |
|  | 6 | 360.33 | 351.77 | 89 | -543.10 |
|  | 5 | 204.88 | 197.68 | 65 | -455.88 |
|  | 3 | 125.39 | 131.36 | 47 | -351.22 |
| 01-02 | 12 | 949.90 | 841.43 | 109 | -258.07 |
|  | 6 | 470.26 | 426.90 | 89 | -470.86 |
|  | 5 | 233.38 | 219.82 | 65 | -435.85 |
|  | 3 | 120.53 | 121.45 | 47 | -352.65 |
| 02-03 | 12 | 952.57 | 873.38 | 109 | -228.43 |
|  | 6 | 434.87 | 428.31 | 89 | -471.33 |
|  | 5 | 244.36 | 235.23 | 65 | -421.81 |
|  | 3 | 113.92 | 115.33 | 47 | -359.76 |

* All p-values are lower than 0.0001.

**Table 9:** Goodness-of-fit statistics for quasi-independence model: professional status

| Panels | Number of categories | $X^2$* | $L^2$* | Degrees of freedom | BIC index |
|--------|--------------|---------|---------|----------|-----------|
| 93-94 | 11 | 3,941.76 | 3,452.42 | 89 | 2,549.77 |
|       | 6  | 1,716.59 | 1,300.78 | 75 | 540.12 |
|       | 2  | 252.24 | 240.00 | 39 | -155.53 |
| 94-95 | 11 | 3,248.49 | 2,765.17 | 89 | 1,877.40 |
|       | 6  | 1,332.34 | 981.83 | 75 | 233.71 |
|       | 2  | 180.09 | 177.74 | 39 | -211.28 |
| 95-96 | 11 | 3,816.64 | 3,182.43 | 89 | 2,281.33 |
|       | 6  | 1,532.71 | 1,143.73 | 75 | 384.38 |
|       | 2  | 210.07 | 207.18 | 39 | -187.68 |
| 96-97 | 11 | 3,496.75 | 3,051.29 | 89 | 2,150.86 |
|       | 6  | 1,346.41 | 1,047.90 | 75 | 289.11 |
|       | 2  | 202.00 | 209.37 | 39 | -185.20 |
| 97-98 | 11 | 3,502.93 | 3,035.06 | 89 | 2,136.46 |
|       | 6  | 1,229.64 | 947.31 | 75 | 190.06 |
|       | 2  | 129.44 | 129.67 | 39 | -264.10 |
| 98-99 | 11 | 3,525.77 | 3,077.31 | 89 | 2,180.46 |
|       | 6  | 1,433.62 | 1,047.19 | 75 | 291.42 |
|       | 2  | 201.73 | 218.96 | 39 | -174.04 |
| 99-00 | 11 | 3,469.02 | 3,103.43 | 89 | 2,205.16 |
|       | 6  | 1,496.51 | 1,134.56 | 75 | 377.59 |
|       | 2  | 198.24 | 208.87 | 39 | -184.75 |
| 00-01 | 11 | 3,509.98 | 2,986.26 | 89 | 2,091.38 |
|       | 6  | 1,280.38 | 958.92 | 75 | 204.81 |
|       | 2  | 172.38 | 180.34 | 39 | -211.80 |
| 01-02 | 11 | 3,454.76 | 3,029.94 | 89 | 2,132.18 |
|       | 6  | 1,168.10 | 890.99 | 75 | 134.45 |
|       | 2  | 169.99 | 175.73 | 39 | -217.67 |
| 02-03 | 11 | 3,439.02 | 3,047.52 | 89 | 2,147.87 |
|       | 6  | 1,324.17 | 990.89 | 75 | 25.55 |
|       | 2  | 186.34 | 161.28 | 39 | -232.94 |

* All p-values are lower than 0.0001.

are due to response categories aggregated in different sectors (Agriculture, Industrial sector, Services) in the two occasions, we note that it underestimates the association between the following couples of categories: Agriculture and Public Administration, defence and compulsory social security; Manufacturing and Wholesale and retail trade; Construction and Professional and support service activities. Instead, the model overestimates association between Manufacturing and Public Administration, defence and compulsory social security; Construction and Wholesale and retail trade; and Professional and support service activities and Agriculture. It is hard to attribute these results only to inconsistencies in responses given one year apart; it appears more convincing to ascribe them to non- random measurement error affecting responses in each wave of the survey.

In the case of professional status, the log-linear model of quasi-independence was estimated so to reproduce the 11-, 6- and 2-category classifications. Table 9 is similar to Table 8, and lists the results of model fitting for the classifications of professional status for the 10 panels. Also for this variable, the hypothesis of quasi-independence is always rejected, for all panels and at all levels of aggregation, indicating that one-year-apart responses show non-random association even at the maximum level of aggregation (2 categories).

The BIC index always reaches its minimum value with the 2-category classification. This indicates that this is the level of aggregation which fits the data best, and again confirms the conclusions reached by applying the hierarchical Kappa procedure.

Inspecting estimated residuals of the quasi-independence model implying 2 categories, i.e., testing quasi-independence among inconsistencies which are due to response categories classified as self-employed on one occasion and as employee on the other, we note that it underestimates the association between the following pairs of categories: Manager and Entrepreneur, Executive and Professional, Clerk and Professional. In this case, patterns of association appear more reasonable and easier to interpret than in the case of industry. The results show that, in fact, residual association tends to concentrate among the highest classes of both employees and self-employed. Usually those in high positions, even if they are employees, enjoy more flexible working conditions, so that they sometimes confuse themselves with self-employed workers. Another sensible explanation is that in Italy, as in other European countries, the standard dichotomising of workers into Employee/Self-employed has recently become too rigid, and is unable to cope with the growth of non-standard forms of employment (see, e.g., Burchell, Deakin and Honey, 1999).

### 3.4. Testing a set of different classifications jointly by professional status and industry

It is of some interest to evaluate what happens in terms of inconsistencies of one-year-apart responses, when we specify joint classifications by professional status and industry, starting from one with 13 classes: Self-employed on the one hand, and employees divided by the 12-category classification of industry on the other.

The aggregation process for the joint classifications is documented in Table 10. It moves from the 13-category classification just mentioned to 7-, 6- and 4-category classifications, respectively, obtained considering Self-employed workers systematically in one category and aggregating employees by industry in 6, 5 and 3 categories, consistently with the strategy for collapsing classes previously used for that variable. In addition to this 4-category classification, denoted classification (a) (Self-employed, Employee in agriculture, Employee in the industrial sector, and Employee in services), we add an alternative 4-category classification recently introduced, on heuristic grounds, by Trivellato *et al.* (2005), which we denote classification (b) (with the four categories given by Self-employed, Employee in agriculture, Employee in industrial sector and private services, and Employee in public administration and social services). Simply stated, the motivation for this alternative classification is the following: the distinction between

**Table 10:** Scheme of categories aggregation process: joint classification by occupational status and industry

| *13 categories* | *7 categories* | *6 categories* | *4 categories: classification (a)* | *4 categories: classification (b)* |
|---|---|---|---|---|
| Self-employed | Self-employed | Self-employed | Self-employed | Self-employed |
| *Employee in:* Agriculture | *Employee in:* Agriculture | *Employee in:* Agriculture | *Employee in:* Agriculture | *Employee in:* Agriculture |
| Mining and raw material extraction | Manufacturing and mining | Manufacturing and mining | Industrial sector | Industrial sector and private services |
| Manufacturing | | | | |
| Construction | Construction | Construction | | |
| Wholesale and retail trade | Wholesale and retail trade | Wholesale and retail trade | Services | |
| Accommodation and food services | Services | Other activities | | |
| Transportation and communications | | | | |
| Financial and real estate activities | | | | |
| Professional and support service activities | | | | |
| Public Administration, defence and compulsory social security | Public Administration | | | Public Administration and social services |
| Education, health and other services | | | | |
| Other public, social and personal service activities | | | | |

employees in (basically) non-market services on the one hand and in the industrial sector and private services on the other is perceived by respondents as more clear than the traditional one which contrasts employees in the industrial sector *vs*. employees in services. The main reason is possibly the extensive process of outsourcing in the industrial sector, which has made the traditional distinction blurred.

Table 11 lists the values of hierarchical Kappa coefficients calculated with these five sets of weights on the 10 panels, and the results of the tests performed on their

**Table 11:** Hierarchical Kappa coefficients and Wald tests: joint classifications by professional status and industry

| Panels | Kappa coefficients | | | | | Wald tests | | |
|---|---|---|---|---|---|---|---|---|
| | 13 categories | 7 categories | 6 categories | 4 categories: classif. (a) | 4 categories: classif. (b) | 7 *vs*. 13 | 6 *vs*. 7 | 4 class. (a) *vs*. 6 |
| 93-94 | 0.8713 | 0.8874 | 0.8972 | 0.9066 | 0.9083 | 242.54*** | 108.54*** | 88.87*** |
| 94-95 | 0.8866 | 0.9020 | 0.9116 | 0.9205 | 0.9195 | 198.08*** | 93.26*** | 73.29*** |
| 95-96 | 0.8822 | 0.8977 | 0.9058 | 0.9123 | 0.9170 | 231.98*** | 84.40*** | 50.84*** |
| 96-97 | 0.8961 | 0.9096 | 0.9148 | 0.9213 | 0.9263 | 198.07*** | 43.66*** | 53.38*** |
| 97-98 | 0.8957 | 0.9046 | 0.9121 | 0.9189 | 0.9224 | 103.14*** | 69.91*** | 54.33*** |
| 98-99 | 0.8881 | 0.9003 | 0.9075 | 0.9143 | 0.9221 | 156.82*** | 63.94*** | 51.69*** |
| 99-00 | 0.8834 | 0.8969 | 0.9036 | 0.9116 | 0.9178 | 183.26*** | 55.30*** | 63.87*** |
| 00-01 | 0.8814 | 0.8942 | 0.8989 | 0.9058 | 0.9162 | 160.76*** | 28.69 | 45.10*** |
| 01-02 | 0.8905 | 0.9014 | 0.9061 | 0.9125 | 0.9234 | 137.23*** | 31.17*** | 43.51*** |
| 02-03 | 0.8945 | 0.9074 | 0.9132 | 0.9181 | 0.9255 | 181.95*** | 14.05*** | 29.19*** |

\* Significant at α=0.1, \*\* significant at α=0.05, \*\*\* significant at α=0.01.

differences. Switching from 13 to 7 categories significantly improves agreement among responses in all panels; reducing categories to 6 and then to 4-classification (a), further significantly increases agreement.

Lastly, we compare the alternative 4-category classification (b) with the standard 4-category classification (a). Also in the case of classification (b) switching from 13 to 4 categories significantly increases agreement among responses in all panels. In addition, the 4-category classification (b) has a higher (and statistically significant) level of agreement in 9 (8) out of 10 panels. The overall $\chi^2$, with 10 degrees of freedom, is 145.05, with a *p*-value close to zero, and definitely confirms that the latter 4-category classification is superior to the former one.

## 4. Conclusions

The focus of this paper is on inconsistencies in job characteristics reported in one-year-apart independent waves of the QLFS by workers continuously employed and who did not change job. Transition matrices by professional status (collected with 11 categories) and industry (recoded into 12 categories) show a significant percentage of frequencies outside the main diagonal.

Aggregating categories improves agreement, as the application of the hierarchical Kappa procedure clearly demonstrates. For professional status the best level of aggregation is the binary one: Employee/Self-employed. For industry two classifications minimise inconsistencies: with 5 classes (Agriculture, Manufacturing and mining, Construction, Wholesale and trade, Other activities) and 6 classes (with a split of Other activities in Services and Public Administration), respectively. In the case of a joint classification by professional status and industry, the best level of aggregation is given by the 4-category classification recently advocated by Trivellato *et. al.* (2005), which distinguishes Self-

employed, Employee in agriculture, Employee in industrial sector and private services, and Employee in public administration and social services.

Inspection of estimated residuals of the log-linear model of quasi-independence – which does not fit the data even at the maximum level of aggregation – suggests that even cross-section information is affected by non-random measurement error, since not all residual association can be explained by inconsistencies among responses perceived as similar by respondents.

Abundant literature (Mathiowetz and McGonagle, 2000; Sala and Lynn, 2006; Lynn, Jäckle, Jenkins and Sala, 2006, among others) shows that dependent interviewing results in lower levels of observed changes in job characteristics collected in two subsequent interviews, compared with changes observed with independent interviewing. It also reveals that this reduction in observed changes coincides with reduction of measurement error, since it is particularly pronounced among workers who do not change job.

The partly disappointing results of our analyses on 1993-2003 two-wave panel data from the QLFS, coupled with evidence from the literature on dependent interviewing, highlight the importance of innovations with the new Continuous LFS, in operation in Italy from 2004. It introduced a definitely better questionnaire and a CAPI-CAPI mode of interview from the start, and is progressively extending dependent interviewing to several sections of the questionnaire, including the one on job characteristics.

Finally, our results bear some significance from the perspective of cross-country studies of job mobility (see, e.g, the recent report to the European Commission by Andersen *et al.*, 2008). It would be improper to extend inferences from a case-study limited to just one country. However, our results about inconsistencies in reported employment characteristics suggest that indicators of job mobility, especially of occupational mobility, based on harmonized household surveys – such as the European Labour Force Survey, the European Community Household Panel and the European Union Statistics on Income and Living Conditions – might be affected by sizable measurement error, possibly varying across countries. Explicit consideration of such measurement error and its effects seems to be essential in order to get reliable (though eventually less detailed) indicators of job mobility.

## References

Andersen T., J. Henrik Haahr, M. Eggert Hansen and M. Holm-Pedersen (2008), *Job mobility in the European Union: Optimising its social and economic benefits. Final report*, Report prepared under contract to the European Commission, Directorate General for Employment, Social Affairs and Equal Opportunities, Aarhus, Danish Technological Institute [http://ec.europa.eu/social/BlobServlet?docId=514&langId=en].

de Angelini A. and A. Giraldo (2003), *La mobilità dei lavoratori nel Veneto. Confronto fra misure su dati RTFL e su dati NETLABOR*, PRIN Research Project "Dinamiche e persistenze nel mercato del lavoro italiano ed effetti di politiche", Working Paper n. 61, Padova, University of Padova, Statistics Department.

Barnhart H.X. and J.M. Williamson (2002), "Weighted least-squares approach for comparing correlated Kappa", *Biometrics*, 58: 1012-1019.

Biemer P.P. (2004), "The twelfth Morris Hansen Lecture simple response variance: Then and now", *Journal of Official Statistics*, 20: 417-439.

Bound J., C. Brown and N. Mathiowetz (2001), "Measurement error in survey data", in J.J. Heckman and E. Leamer (eds.), *Handbook of Econometrics. Volume 5*, Amsterdam: Elsevier Science.

Burchell B., S. Deakin and S. Honey (1999), *The employment status of individuals in non-standard employment*, EMAR Publications No. 6, London: Department of Trade and Industry.

Cohen J. (1960), "A coefficient of agreement for nominal tables", *Educational and Psychological Measurement*, 20: 37-46.

Cohen J. (1968), "Weighted kappa: Nominal scale agreement with provision for scale disagreement or partial credit", *Psychological Bulletin*, 70: 213-220.

Eurostat (2000), "Commission Regulation (EC) No 1575/2000 of 19 July 2000 implementing Council Regulation (EC) No 577/98 on the organisation of a labour force sample survey in the Community concerning the codification to be used for data transmission from 2001 onwards", *Official Journal of the European Community*, 43 (L181), 20.7.2000: 16-34.

Fallick B. and C.A. Fleischman (2004), *Employer-to-employer flows in the U.S. labour market: The complete picture of gross worker flows*, Finance and Economics Discussion Series 2004-34, Washington D.C.: Board of Governors of the Federal Reserve System (U.S.).

Fleiss J.L. (1981), *Statistical methods for rates and proportions*, New York: Wiley.

Gandolfo M. and P. Gennari (2000), "Il lavoro sommerso, è possibile rilevarlo? In Società Italiana di Statistica", Atti della XL Riunione Scientifica. Firenze 26-28 April 2000, University of Florence, Statistics Department.

Goodman L.A. (1968), "The analysis of cross-classified data: independence, quasi-independence, and interaction in contingency tables", *Journal of the American Statistical Association*, 63: 1019-1131.

Hagenaars J.A. (1990), *Categorical longitudinal data*, Newbury Park: Sage.

Hansen M., W.N. Hurwitz and L. Pritzker (1964), "The estimation and interpretation of gross differences and the simple response variance", in C.R. Rao (ed.), *Contributions to Statistics* (presented to C. Mahalanobis on the occasion of his 70th birthday), Calcutta: Statistical Publishing Society.

Hauser R.M. and M.P. Massagli (1983), "Some models of agreement and disagreement in repeated measurements of occupation", *Demography*, 20: 449-460.

Koch G.G., J.R. Landis, J.L. Freeman, D.H. Freeman and R.G. Lehnen (1977), "A general methodology for the analysis of experiments with repeated measurements of categorical data", *Biometrics*, 33: 133-158.

ILO (1993), *Resolution concerning the International Classification of Status in Employment (ICSE)*, Adopted by the Fifteenth International Conference of Labour Statisticians (28 January 1993), Geneva [http://www.ilo.org/wcmsp5/groups/public/---dgreports/---integration/---stat/documents/normativeinstrument/wcms_087562.pdf].

ILO (2007), *Resolution concerning updating the International Standard Classification of Occupation*, adopted by the Tripartite Meeting of Experts on Labour Statistics on Updating the International Standard Classification of Occupation (ISCO), Geneva, 7 December 2007 [http://www.ilo.org/public/english/bureau/stat/isco/docs/resol08.pdf].

Istat (2003), *Rilevazione trimestrale sulle forze di lavoro. Tracciato record del file standard longitudinale (Aprile 2003-Aprile 2002)*, Roma (mimeo).

Istat (2004), *La nuova rilevazione sulle forze di lavoro - Contenuti, metodologie, organizzazione*, Roma.

Landis J.R. and G.G. Koch (1977), "The measurement of observer agreement for categorical data", *Biometrics*, 33: 159-174.

Leombruni R. and R. Quaranta (2002), "Mobilità dei lavoratori in Italia, 1985-1996", in B. Contini (ed.), *Osservatorio sulla mobilità del lavoro in Italia*, Bologna: Il Mulino.

Leombruni R. and R. Quaranta (2005), "Eppure si muoveva già. Una breve storia della mobilità dei lavoratori in Italia", in B. Contini and U. Trivellato (eds.), *Eppur si muove. Dinamiche e persistenze nel mercato del lavoro italiano*, Bologna: Il Mulino.

Lynn P., A. Jäckle, S.P. Jenkins and E. Sala (2006), "The effects of dependent interviewing on responses to questions on income sources", *Journal of Official Statistics*, 22: 357-384.

Mathiowetz N. (1992), "Errors in reports of occupation", *Public Opinion Quarterly*, 56: 352-355.

Mathiowetz N. and A. McGonagle (2000), "An assessment of the current state of dependent interviewing in household surveys", *Journal of Official Statistics*, 16: 401-418.

Mellow W. and H. Sider (1983), "Accuracy of response in labor market surveys: Evidence and implications", *Journal of Labor Economics*, 1: 331-344.

McCall B.P. (1990), "Occupational matching: A test of sorts", *Journal of Political Economy*, 89: 45-69.

Miller R.A. (1984), "Job matching and occupation choice", *Journal of Political Economy*, 92: 1086-1120.

Neal D. (1999), "The complexity of job mobility among young men", *Journal of Labour Economics*, 17: 237-261.

Sala E. and P. Lynn (2006), "Measuring change in employment characteristics: The effects of dependent interviewing", *International Journal of Public Opinion Research*, 18, 500-509.

Shimer R. (2005), "The cyclicality of hires, separations, and job-to-job transitions", *Federal Reserve Bank of St Louis Review*, 87: 493-597.

Steven D.J. and J. Haltiwanger (1999), "Measuring gross worker and job flows," in J. Haltiwanger, M. Manser and R. Topel (eds.), *Labor Statistics Measurement Issues*, Chicago: University of Chicago Press.

Trivellato U. (1997), "Le misure della partecipazione al lavoro nel quadro comunitario", in L. Frey (ed.), *Le informazioni sul lavoro in Italia: significato e limiti delle informazioni provenienti da indagini sulle famiglie, Quaderni di Economia del Lavoro n. 59*, Milano: Franco Angeli.

Trivellato U., A. Paggiaro, R, Leombruni and S. Rosati (2005), "La dinamica recente della mobilità dei lavoratori, 1998-2003", in B. Contini and U. Trivellato (eds.), *Eppur si muove. Dinamiche e persistenze nel mercato del lavoro italiano*, Bologna: Il Mulino.