STATISTICA, anno LXXII, n. 1, 2012

INCONSISTENCIES IN REPORTED EMPLOYMENT CHARACTERISTICS AMONG EMPLOYED STAYERS (*)

F. Bassi, A. Padoan, U. Trivellato

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 (Davis and Haltiwanger, 1998; Fallick and Fleischman, 2004; Shimer, 2005, among many others). The complexity of job mobility also demands analysing the kind of job changes that workers experience 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, at least when information is obtained at two points in time with independent interviews (Bound *et al.*,, 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

(*) Research for this paper was supported by grants 2003139334 and 2005131989 from the Italian Ministry for Education, University and Research, under the PRIN Programme. Individual anonymised data from the Italian Labour Force Survey were kindly provided by Istat (Italian statistical agency), under a research agreement with the Department of Statistics, University of Padova. Earlier versions were presented at the European Conference on Quality in Official Statistics 2008, Rome, July 2008, the 3rd Conference of ESRA, Warsaw, June 2009, and the Statistics Canada's International Symposium "Longitudinal Surveys: from Design to Analysis", Gatineau, October 2009. We are grateful to Guido Masarotto and Adriano Paggiaro, conference participants and especially a referee for useful comments and suggestions. The usual disclaimer applies.

scheme, is higher than that resulting when reports are categorised according to the more detailed three-digit classification (Mellow and Sider, 1983; Mathiowetz, 1992). Two recent contributions on the topic are by Kambourov and Manovskii (2008), who develop a method for controlling for measurement error in industry and occupation coding in the Panel Study of Income Dynamics, based on retrospective information from the same survey, and Isaoglu (2010) who investigates on several measurement errors in occupational classifications in the German Socio-Economic Panel, presents a correction method based on job or labour market status, and analyzes its impact on patterns of occupational mobility.

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 with independent interviews by the Italian Quarterly Labour Force Survey (QLFS) in the period from April 1993 to April 2003.

The QLFS is a fundamental source for the analysis of short-term dynamics and persistence in the Italian labour market. 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. The purpose of the paper is twofold. (i) We present a sensible strategy for analysing this kind of inconsistencies in LFSs with a panel component, and document their pattern and amount for the Italian QLFS. (ii) We claim that results should be of interest both to official statistical agencies and to analysts of labour market micro-dynamics. The former could provide better guidelines for a judicious use of the data, and get clues for improving the data collection process. The latter should be motivated to carry out their studies taking this type of measurement error into due account.

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. We then expand our analysis in three directions: (i) we test whether the consistency of repeated information 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.

[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 Occupations (ISCO), and by "status of employment", as defined by the International Classification of Status of Employment (ICSE) (see ILO, 2008). Following Eurostat (2000), from now on we will call it "professional status".

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 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. Besides, inspection of estimated residuals suggests that even cross-section information is affected by measurement error.

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. Interestingly enough, the latter classification turns out to be almost uniformly better than the former 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[2]. 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 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). In-

[2] The description of the survey given here applies until 2003. In 2004, the survey was substantially redesigned (Istat, 2004).

formation on industry is collected with an open-ended question and coded by the interviewer according to the ATECO2002[3] classification. As already noted, Istat (2003) warns that these two variables may not be reliable if used at their maximum degree of detail, and suggests using the binary classification, Employee/Self-employed, for professional status, and a 12-category classification, corresponding to the two-digit ATECO2002, for industry[4].

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 240,657 sample units, around 24,000 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 – by day, month and year – when answering the question: "When did you start working with the current employer or in the current self-employment?"[5].

We assume that no or negligible errors are made in reporting dates, and that, among yearly employment stayers who did not change employer or current self-employment, genuine levels of change in industry and/or professional status are likely to be very low (Sala and Lynn, 2006). Thus, 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 open-ended question used to collect information.

2.2. *Methods and design of analyses*

Transition matrices among job characteristics declared one year apart provide the basic information for quantifying inconsistencies. As an example, Table 1 reports the transition matrix by industry 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.

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.

---

[3] ATECO 2002 is identical to NACE Rev. 1.1 at four-digit level.

[4] The 12 categories for industry are: Agriculture; Mining and raw material extraction; Manufacturing; Construction; Wholesale and retail trade; Accommodation and food services; Transportation and communications; Financial and real estate activities; Professional and support service activities; Public administration, defence and compulsory social security; Education, health and other social services; Other public, social and personal service activities.

[5] 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.

A simple indicator of disagreement is the percentage of frequencies outside the main diagonal (or gross difference rate, as Hansen *et al.*, 1964, call it).

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[6].

TABLE 1

Transition matrix by industry, April 1993 to April 1994

	1994											
1993	Agric.	Mining	Manuf.	Constr.	Wholes.	Accom.	Transp.	Finance	Profess.	P.A.	Education	Other
Agric.	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
Manuf.	16	28	5315	87	179	15	41	11	43	29	37	36
Constr.	12	12	90	1,721	28	0	16	6	27	17	12	21
Wholes.	24	6	171	55	3,648	40	31	6	22	19	28	34
Accom.	2	1	6	5	26	705	10	0	0	6	15	14
Transp.	11	2	47	22	31	4	1,306	10	4	44	11	21
Finance	2	5	17	8	21	2	10	777	12	12	11	10
Profess.	2	2	39	39	32	3	9	23	814	22	22	65
P.A.	23	11	37	22	25	10	48	13	12	2,098	164	27
Education	7	2	25	10	28	13	10	10	11	108	3,188	38
Other	9	5	25	14	39	19	25	8	36	40	58	938

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 re-

[6] 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 are 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%). Unfortunately, information on self/proxy respondents was not included in the micro-data sets made available to us by Istat. For these reasons, we prefer Cohen's Kappa to the intraclass correlation coefficient. It is worth adding that a strand of the literature offers insights into theoretical and practical development of Cohen's Kappa (see Gwet, 2002, Sim and Wright, 2005, and Gwet, 2008, among others). It points out that Cohen's Kappa has some limitations; for instance it is influenced by factors such as trait prevalence, and may yield unexpected results in the presence of high agreement. We will deal with the issues complementing the evidence provided by Kappa coefficients with analyses based on log-linear models of quasi-independence: a route that appears to use more information than resorting to alternative coefficients of agreement.

sponse 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 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 $H_o : K_2 = K_1$ *vs.* $H_1 : K_2 > K_1$ 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. They take 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).

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 of yearly employed stayers who did not change job, as described in Section 2.1, Table 2 contains the values of the descriptive statistics of inconsistency: the percentage of frequencies 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[7]. 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.

TABLE 2

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.

In order to interpret the values of Cohen's Kappa in empirical studies two scales are mainly used, proposed by Fleiss (1981) and Landis and Koch (1977), respectively.

As shown in Table 2, for both variables – industry and professional status – the values of Cohen's Kappa are fairly high, larger than 0.8 for professional status and close to 0.9 for industry. 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 coefficient equal to or very close to 1, which does not seem to be the case[8]. 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.

---

[7] As already mentioned, this result can also be found in other studies. However, in our case-study it is not trivial, 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).

[8] 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.

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 3 and 5, was based both on judgment and on evidence on the distribution and size of inconsistencies across categories documented by the descriptive analyses.

TABLE 3

Scheme of categories aggregation process: industry

<table>
<tr><th>12 categories</th><th>6 categories</th><th>5 categories</th><th>3 categories</th></tr>
<tr><td>Agriculture</td><td>Agriculture</td><td>Agriculture</td><td>Agriculture</td></tr>
<tr><td>Mining and raw material extraction</td><td rowspan="2">Manufacturing and mining</td><td rowspan="2">Manufacturing and mining</td><td rowspan="3">Industrial sector</td></tr>
<tr><td>Manufacturing</td></tr>
<tr><td>Construction</td><td>Construction</td><td>Construction</td></tr>
<tr><td>Wholesale and retail trade</td><td>Wholesale and retail trade</td><td>Wholesale and retail trade</td><td rowspan="8">Services</td></tr>
<tr><td>Accommodation and food services</td><td rowspan="4">Services</td><td rowspan="7">Other activities</td></tr>
<tr><td>Transportation and communications</td></tr>
<tr><td>Financial and real estate activities</td></tr>
<tr><td>Professional and support service activities</td></tr>
<tr><td>Public Administration, defence and compulsory social security</td><td rowspan="3">Public Administration</td></tr>
<tr><td>Education, health and other services</td></tr>
<tr><td>Other public, social and personal service activities</td></tr>
</table>

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. The third set of weights corresponds to a 5-category classification, derived from the previous one. The last set of weights implies the usual 3-category classification: Agriculture, Industrial sector, and Services (see Table 3)[9].

[9] An example might be useful. Let us start from the transition matrix reported in Table 1, i.e., from a classification of industry on 12 categories, and decide to switch from 12 to 6 categories. Calculating the appropriate weighted Kappa implies to set equal to 1 the weights of the following cells: (1,1), (1,2), (2,1), (2,2), (3,3), (4,4), (5,5), (6,7), (6,8), (6,9), (7,8), (7,9), (8,9), (7,6), (8,6), (9,6), (8,7), (9,7), (9,8), (6,), (6,7), (8,8), (9,9), (10,11), (10,12), (11,12),(11,10),(12,10), (12,11), (10,10), (11,11), (12,12).

Table 4 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. 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.

TABLE 4

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=0.05, ** significant at α=0.01.

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. The last set of weights corresponds to the binary classification recommended by Istat: Employee and Self-employed (see Table 5).

TABLE 5

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	

Table 6 refers to the three sets of classifications by professional status, and is similar to Table 4. Switching from 11 to 6 categories significantly improves agreement among responses in all panels; 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.

TABLE 6

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.

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

The log-linear model of quasi-independence was estimated so to reproduce the 12-, 6-, 5- and 3-category classifications for industry, and the 11-, 6- and 2-category classifications for occupational status. Table 7 lists the results of model fitting (Pearson X^2 and BIC index) for the 10 panels.

For industry, 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 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.

TABLE 7

Goodness-of-fit statistics for quasi-independence model: industry and professional status

	Industry				*Professional status*			
Panels	No. of categories	X^2*	Degrees of freedom	BIC index	No. of categories	X^2*	Degrees of freedom	BIC index
93-94	12	1,310.95	109	74.87	11	3,941.76	89	2,549.77
	6	651.81	89	-306.12	6	1,716.59	75	540.12
	5	350.64	65	-350.27	2	252.24	39	-155.53
	3	202.72	47	-286.42				
94-95	12	1,117.23	109	-93.73	11	3,248.49	89	1,877.40
	6	557.06	89	-373.20	6	1,332.34	75	233.71
	5	234.53	65	-428.75	2	180.09	39	-211.28
	3	111.99	47	-360.52				
95-96	12	1,073.58	109	-138.19	11	3,816.64	89	2,281.33
	6	520.16	89	-399.61	6	1,532.71	75	384.38
	5	253.15	65	-417.77	2	210.07	39	-187.68
	3	145.81	47	-330.67				
96-97	12	1,038.38	109	-206.29	11	3,496.75	89	2,150.86
	6	411.06	89	-506.63	6	1,346.41	75	289.11
	5	220.96	65	-445.05	2	202.00	39	-185.20
	3	117.53	47	-355.74				
97-98	12	776.94	109	-407.74	11	3,502.93	89	2,136.46
	6	359.91	89	-533.31	6	1,229.64	75	190.06
	5	240.31	65	-426.97	2	129.44	39	-264.10
	3	111.58	47	-362.59				
98-99	12	990.77	109	-210.85	11	3,525.77	89	2,180.46
	6	511.34	89	-407.79	6	1,433.62	75	291.42
	5	277.61	65	-381.76	2	201.73	39	-174.04
	3	138.15	47	-331.40				
99-00	12	1,087.13	109	-102.56	11	3,469.02	89	2,205.16
	6	542.33	89	-374.25	6	1,496.51	75	377.59
	5	284.80	65	-382.92	2	198.24	39	-184.75
	3	183.87	47	-299.86				
00-01	12	976.55	109	-234.43	11	3,509.98	89	2,091.38
	6	360.33	89	-543.10	6	1,280.38	75	204.81
	5	204.88	65	-455.88	2	172.38	39	-211.80
	3	125.39	47	-351.22				
01-02	12	949.90	109	-258.07	11	3,454.76	89	2,132.18
	6	470.26	89	-470.86	6	1,168.10	75	134.45
	5	233.38	65	-435.85	2	169.99	39	-217.67
	3	120.53	47	-352.65				
02-03	12	952.57	109	-228.43	11	3,439.02	89	2,147.87
	6	434.87	89	-471.33	6	1,324.17	75	25.55
	5	244.36	65	-421.81	2	186.34	39	-232.94
	3	113.92	47	-359.76				

* All p-values are lower than 0.0001.

Also for professional status the hypothesis of quasi-independence is always rejected, for all panels and at all levels of aggregation, pointing to non-random association of one-year-apart responses even at the maximum level of aggregation (2 categories).

The BIC index always reaches its minimum value with the 2-category classification. This 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. A 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 *et al.*, 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 8. 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 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 9 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 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 χ^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.

TABLE 8

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

<table>
<tr><th>13 categories</th><th>7 categories</th><th>6 categories</th><th>4 categories: classification (a)</th><th>4 categories: classification (b)</th></tr>
<tr><td>Self-employed</td><td>Self-employed</td><td>Self-employed</td><td>Self-employed</td><td>Self-employed</td></tr>
<tr><td>Employee in:</td><td>Employee in:</td><td>Employee in:</td><td>Employee in:</td><td>Employee in:</td></tr>
<tr><td>Agriculture</td><td>Agriculture</td><td>Agriculture</td><td>Agriculture</td><td>Agriculture</td></tr>
<tr><td>Mining and raw material extraction</td><td rowspan="2">Manufacturing and mining</td><td rowspan="2">Manufacturing and mining</td><td rowspan="3">Industrial sector</td><td rowspan="8">Industrial sector and private services</td></tr>
<tr><td>Manufacturing</td></tr>
<tr><td>Construction</td><td>Construction</td><td>Construction</td></tr>
<tr><td>Wholesale and retail trade</td><td>Wholesale and retail trade</td><td>Wholesale and retail trade</td><td rowspan="8">Services</td></tr>
<tr><td>Accommodation and food services</td><td rowspan="4">Services</td><td rowspan="7">Other activities</td></tr>
<tr><td>Transportation and communications</td></tr>
<tr><td>Financial and real estate activities</td></tr>
<tr><td>Professional and support service activities</td></tr>
<tr><td>Public Administration, defence and compulsory social security</td><td rowspan="3">Public Administration</td><td rowspan="3">Public Administration and social services</td></tr>
<tr><td>Education, health and other services</td></tr>
<tr><td>Other public, social and personal service activities</td></tr>
</table>

TABLE 9

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	4 class. (b) *vs.* 13
93-94	0.8713	0.8874	0.8972	0.9066	0.9083	242.54**	108.54**	88.87**	492.59**
94-95	0.8866	0.9020	0.9116	0.9205	0.9195	198.08**	93.26**	73.29**	370.86**
95-96	0.8822	0.8977	0.9058	0.9123	0.9170	231.98**	84.40**	50.84**	466.25**
96-97	0.8961	0.9096	0.9148	0.9213	0.9263	198.07**	43.66**	53.38**	393.49**
97-98	0.8957	0.9046	0.9121	0.9189	0.9224	103.14**	69.91**	54.33**	315.62**
98-99	0.8881	0.9003	0.9075	0.9143	0.9221	156.82**	63.94**	51.69**	439.54**
99-00	0.8834	0.8969	0.9036	0.9116	0.9178	183.26**	55.30**	63.87**	438.78**
00-01	0.8814	0.8942	0.8989	0.9058	0.9162	160.76**	28.69	45.10**	422.11**
01-02	0.8905	0.9014	0.9061	0.9125	0.9234	137.23**	31.17**	43.51**	419.25**
02-03	0.8945	0.9074	0.9132	0.9181	0.9255	181.95**	14.05**	29.19**	390.21**

** significant at α=0.01.

4. CONCLUSIONS

The focus of this paper is on inconsistencies in job characteristics reported in one-year-apart independent interviews 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 significantly 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[10].

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 *et al.*, 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 highlight the importance of innovations brought in with the new Continuous LFS (CLFS), in operation from 2004 (Istat, 2004). Among them, the new questionnaire and the mixed CAPI-CATI mode of interview are especially important. Clearly, the issue of inconsistencies in reported employment characteristics that affect the CLFS is now on the agenda.

Nonetheless, the results presented in this paper are of interest for several reasons. First of all, it is worth stressing that the new CLFS collects information on industry and professional status still with the same questions and format used in old QLFS (the only exception is that two categories have been added to the close-form question of professional status). Hence, the methods used and the strategy presented for assessing classification inconsistencies in the QLFS could be fruitfully employed for addressing the same questions with data from the CLFS. Besides, it is reasonable to expect that the patterns of measurement error in industry and professional status revealed by our analyses on the QLFS will affect also the new CLFS, at least in part. Thus, the evidence from our study provides caveats and useful, though rudimentary, clues for handling the information on occupational mobility provided by the CLFS. Finally, our results, coupled with evidence from the literature on dependent interviewing, suggest that improvements on the wording of questions and on dependent interviewing approaches

[10] Researchers focused on substantive issues of occupational mobility may be disappointed by the loss of information resulting from the aggregation procedures presented in this paper. However, there is no real loss. Indeed, the detailed classification they have to give up is misleading: plagued by non-random measurement error, which would inflate dramatically the occupational mobility patterns. The target is to get reliable detailed information on employment characteristics. The only way to reach it rests on improvements in data collection strategies.

(e.g., proactive or reactive) might be crucial for response error reduction in reported employment characteristics. To that purpose, small-scale experiments would be quite useful for assessing the effectiveness of different interrogation strategies and dependent interviewing approaches.

Department of Statistics FRANCESCA BASSI
University of Padova

Office of Statistics, Regione Veneto ALESSANDRA PADOAN

Department of Statistics UGO TRIVELLATO
University of Padova

REFERENCES

H.X. BARNHART, J.M. WILLIAMSON (2002), *Weighted least-squares approach for comparing correlated Kappa*, “Biometrics”, 58, pp. 1012-1019.

J. BOUND, C. BROWN, N. MATHIOWETZ (2001), *Measurement error in survey data*, in J.J. Heckman and E. Leamer (eds.), “Handbook of Econometrics, Volume 5”, Elsevier Science, Amsterdam, pp. 3705-3843.

B. BURCHELL, S. DEAKIN, S. HONEY (1999), *The employment status of individuals in non-standard employment*, EMAR Publications No. 6, Department of Trade and Industry, London.

J. COHEN (1960), *A coefficient of agreement for nominal tables*, “Educational and Psychological Measurement”, 20, pp. 37-46.

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

S.J. DAVIS, J. HALTIWANGER (1998), *Measuring gross worker and job flows*, in J. Haltiwanger, M. Manser and R. Topel (sds.), “Labor Statistics Measurement Issues”, University of Chicago Press, Chicago, pp. 77-119.

EUROSTAT (2000), *Council Regulation (EC) No 1575/2000 of 19 July 2000 implementing Council Regulation (EC) No 577/1998 on the organization of the 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”, L. 181, 20.07.2000, 16-33.

B. FALLICK, 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, Board of Governors of the Federal Reserve System (U.S.), Washington D.C..

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

M. GANDOLFO, P. GENNARI (2000), *Il lavoro sommerso, è possibile rilevarlo?* In Società Italiana di Statistica, *Atti della XL Riunione Scientifica. Firenze 26-28 Apriel 2000*, Università di Firenze, Dipartimento di Statistica G. Parenti.

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

K. GWET (2002), *Inter-rater reliability: Dependency on trait prevalence and marginal homogeneity*, “Statistical Methods for Inter-Rater Reliability Assessment”, 2, pp. 1-9.

K. GWET (2008), *Computing inter-rater reliability and its variance in the presence of high agreement*, “British Journal of Mathematical and Statistical Psychology”, 2, pp. 1-9.

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

M. HANSEN, W.N. HURWITZ, L. PRITZKER (1964), *The estimation and interpretation of gross differences* *and the simple response variance*, in C. Rao (ed.), "Contributions to Statistics" (presented to C. Mahalanobis on the occasion of his 70th birthday), Statistical Publishing Society, Calcutta, pp. 111-136.

G. KAMBOUROV, I. MANOVSKII (2008), *Rising occupation and industry mobility in the United States: 1968-1997*, "International Economic Review", 49, pp.41-79.

G.G. KOCH, J.R. LANDIS, J.L. FREEMAN, D.H. FREEMAN, R.G. LEHNEN (1977), *A general methodology for the analysis of experiments with repeated measurements of categorical data*, "Biometrics", 33, pp. 133-158.

ILO (2008). *Resolution Concerning Updating the International Standard Classification of Occupation*, Geneva.

A. ISAOGLU (2010), *Occupational Affiliation Data and Measurement Error in the German Socio-Economic Panel*, SOEP papers 318, DIW, Berlin.

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

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

J.R. LANDIS, G.G. KOCH (1977), *The measurement of observer agreement for categorical data*, "Biometrics", 33, pp. 159-174.

P. LYNN, A. JÄCKLE, S.P. JENKINS, E. SALA (2006), *The effects of dependent interviewing on responses to questions on income sources*, "Journal of Official Statistics", 22, pp. 357-384.

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

N. MATHIOWETZ, A. MCGONAGLE (2000), *An assessment of the current state of dependent interviewing in household surveys*, "Journal of Official Statistics", 16, pp. 401-418.

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

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

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

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

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

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

J. SIM, C.C. WRIGHT (2005), *The Kappa statistics in reliability studies: use, interpretation, and sample size requirements*, "Physical Therapy", 85, pp.257-268.

U. TRIVELLATO (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, Franco Angeli, Milano, pp. 9-34.

U. TRIVELLATO, A. PAGGIARO, R. LEOMBRUNI, 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", Il Mulino, Bologna, pp. 271-323.

SUMMARY

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 jointly by professional status and occupation. 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.