

Worker Mobility from Social Security Registers and Household Surveys: A Comparative Analysis ⁽¹⁾

La Mobilità dei Lavoratori da Fonti Amministrative e da 'Surveys' sulle Famiglie: un'Analisi Comparata

Paolo Baretta and Ugo Trivellato

Dipartimento di Scienze Statistiche, Università di Padova

via C. Battisti 241, 35121 Padova, e-mail: baretta@stat.unipd.it , trivell@stat.unipd.it

Riassunto: In Italia le analisi sulla mobilità dei lavoratori utilizzano prevalentemente gli archivi dell'Inps. Si esamina la possibilità di utilizzare per questo scopo la Rilevazione Trimestrale delle Forze di Lavoro (RTFL). Si identificano e si controllano le principali ragioni di divergenza fra le misure di mobilità fornite dalle due fonti – diversa popolazione di riferimento e diverso piano di osservazione –. Si propone un metodo detto 'dell'abbinamento su tre occasioni', che utilizza dati abbinati di tre rilevazioni della RTFL nell'arco dell'anno, e se ne valutano le *performances*. L'analisi empirica è condotta intorno all'anno 1995. L'evidenza segnala la capacità del metodo, e quindi dei corrispondenti *data set* longitudinali della RTFL, di cogliere larga parte della mobilità dei lavoratori documentata dall'Inps.

Keywords: Administrative data, Employment spells, Household surveys, Worker turnover and reallocation.

1. Introduction

In the last two decades the literature on labour mobility has experienced remarkable developments, as regard both measures of worker and job flows and empirical analyses on the level and patterns of mobility in several developed countries (see, e.g., Davis, Haltiwanger, 1999). Methods and results are interwoven with the data used, either from surveys or from a variety of administrative sources.

In Italy most of the research on worker and job mobility has been carried out on social security (Inps) administrative databases, particularly on an employer-employee linked panel database developed from them by a team of researchers now at LABORatorio R. Revelli, Center from Employment Studies (Contini, 2002). This panel database – INPS in the sequel – provides detailed information, on a monthly basis, about all employment spells of a 1/90 systematic sample of the relevant population of workers. The population consists of those workers who pay compulsory contributions to Inps: essentially, employees in the private non-agricultural sector. Thus its coverage is partial, though large. Besides, INPS is updated irregularly and with delay.

We investigate the possibility of using the Italian Quarterly Labour Force Survey (RTFL) as an additional source for currently measuring worker mobility. In principle, the motivation for using RTFL is twofold: it covers (almost) the whole employment,

⁽¹⁾ Research for this paper has been supported by a MIUR grant to the project "Dynamics and inertia in the Italian labour market and policy evaluation".

while, as just noted, INPS is restricted to dependent employment in the private non-agricultural sector; it is timely, while INPS is available with a lag of 4-5 years. The RTFL suffers, however, from a serious hindrance: the observation plan of the questionnaire provides fragmentary information on work histories. At each survey wave, in addition to current state – employed, unemployed or inactive, the only information collected is on (i) when the current spell of employment begun (year and month), for the employed, or (ii) when the last employment spell ended (year and month again), for the unemployed or inactive with previous working experience.

Our aims are essentially two:

- to identify and, to a large extent, control for the main reasons of discrepancy of worker mobility indicators from the two sources;
- to ascertain if a method can be developed, based on the RTFL, that is capable of largely reconcile the evidence about worker mobility from the two sources.

2. Worker Mobility Indicators and Design of Comparative Analyses

Our approach goes through four main steps.

First, we suggest a refinement to the usual indicators of worker mobility. Let denote by A_i and S_i the number of engagements and separations, respectively, of individual i in a given time interval – typically a calendar year, and by R_i an indicator variable which is 1 if individual i experiences at least one transition within the time interval. It is appropriate to compute the mobility indicators – engagement rate (a), separation rate (s), gross worker turnover (t) and reallocation rate (r) – as the ratio of the total number of relevant transitions (or individuals) to the population at risk, which we denote by N_f . Thus, our favorite worker mobility indicators are the following:

$$a_f = \frac{1}{N_f} \sum_{i=1}^{N_f} A_i, \quad s_f = \frac{1}{N_f} \sum_{i=1}^{N_f} S_i, \quad t_f = a_f + s_f, \quad r_f = \frac{1}{N_f} \sum_{i=1}^{N_f} R_i. \quad (1)$$

Note that, in general, N_f differs from a measure of the stock of employed, such as $N_s = (N_{si} + N_{sf}) / 2$, i.e., the mean stock of employed at the beginning and at the end of the time interval. N_s is commonly used in the literature, which results in the parallel indicators a_s , s_s , t_s and r_s . (If we consider just engagements and separations pertaining to those employment spells which are ongoing at the beginning and at the end of the year, we get the corresponding indicators ${}_w a_s$, ${}_w s_s$, ${}_w t_s$ and ${}_w r_s$.)

Second, we introduce a new approach for extracting as much information as possible on work histories from the RTFL. The basic idea is to exploit the rotating panel design of RTFL – a 2-2-2 one – in order to link the records pertaining to the same individual over subsequent waves within a calendar year. Our ‘three-waves matching method’ – as we call it – does exactly that.

- By means of a record linkage procedure we generate three-waves panels from the sequence of RTFL surveys of a given year, for the occasions January-March-January and January-October-January respectively. We combine current state information and retrospective information from each of the three waves. The resulting observation plan allows one to detect up to two engagements and two separations (and three employment spells) for each individual.

- Furthermore, as the two three-waves panels are disjoint and independent samples from the same population, we aggregate them. We get a pooled sample whose size is roughly doubled, with an appreciable gain in the precision of worker mobility estimates.

Third, in order to carry out an informative comparative analysis on worker mobility measures from the two sources – INPS vs. the RTFL three-waves matching method – we design a procedure that allows us to control for the two crucial factors of divergence: the reference population and the observation plan. To that purpose, we proceed to two interconnected operations.

- (a) We identify a common reference aggregate, thus getting rid of any systematic difference in mobility measures due to differences in population coverage and composition. (We will refer to such aggregate as the ‘standardized population’).
- (b) We simulate the RTFL three-way matching method onto the INPS databases. As a result, we get worker mobility estimates from the two sources comparable also from the point of view of the observation plan.

Based on these operations, we introduce a decomposition of the discrepancy of mobility indicators from the two sources, which focuses on the relative importance of the two main factors of divergence – population coverage and observation plan – and on their overall capacity to account for the discrepancy.

Finally, we complement the point estimates of mobility indicators with 95% bootstrap confidence intervals. They are especially useful for a straightforward assessment of the ‘closeness’ of mobility measures from the two sources, based on the overlapping/non-overlapping of the intervals themselves.

3. Selected Results

The empirical analysis is carried out for the year 1995. Just summary results are presented here (see Table 1); for extended results see Baretta, Trivellato (2003).

Table 1: Worker mobility measures (three-waves matching method; indicators in %)

Reference population, observation plan, and type of mobility measure	Gross worker turnover		Reallocation rate	
	INPS	RTFL	INPS	RTFL
‘Raw’ measures ^a	65.5	17.6	41.0	17.6
‘Optimal’ obs. plan & type of measure ^b	52.4	—	32.8	—
Fully normalized comparisons ^c	46.5	47.8	32.6	32.9
Standardized population only ^d	51.1	47.8	33.0	32.9
Normalized obs. Plan only ^e	47.1	33.4	32.3	21.0

^a Different reference population and different observation plan (for RTFL the plan uses just information from a single cross-section); measures are t_s and r_s (for RTFL we have $t_s \equiv r_s \equiv {}_w t_s \equiv {}_w r_s$).

^b Observation plan of INPS, based on monthly information (panel method); measures are t_f and r_f .

^c Standardized population, three-waves matching method; mean INPS and pooled RTFL; measures are t_f and r_f .

^d Standardized population, different observation plan (the best for each source: panel method for INPS, three-waves matching method for RTFL); pooled RTFL; measures are t_f and r_f .

^e Different population; three-waves matching method; mean INPS and pooled RTFL; measures are t_f and r_f .

Crude, raw analyses on the two sources, which do not take into account any factor of divergence between them and do not make any use of the panel dimension of RTFL, produce mobility measures from INPS and RTFL which are far apart (Table 1, first row).

On the contrary, after accounting for the two main factors of divergence we get estimates of both gross worker turnover and reallocation rate pretty close: 46.5 vs. 47.8% and 32.6 vs. 32.9%, respectively (Table 1, third row). For both indicators the estimates from the two sources do not differ significantly.

Clearly, the most interesting exercise consists of comparing the two sources on the standardized population, but using each of them at its best as regard the observation plan: that based on monthly information – the panel method – for INPS and the three-waves matching method for RTFL. In this case too mobility estimates turn out to be reasonably similar: 51.1 vs. 47.8% for gross worker turnover and 33.0 vs. 32.9 % for the reallocation rate (Table 1, fourth row). The reallocation rate estimates are hardly distinguishable (the corresponding confidence intervals largely overlap). This is not the case for gross worker turnover estimates, whose difference is statistically significant, though substantively moderate: RTFL underestimates the benchmark turnover resulting from INPS by 3.3 percentage points, around 6.5% in relative terms.

4. Concluding Remarks

The overall evidence is that the three-waves matching method captures large part of the worker mobility documented by INPS: entirely as worker reallocation measures are concerned; with a significant, but moderate underestimation with respect to gross worker turnover.

The tentative conclusion is that longitudinal datasets from RTFL might be used for satisfactorily estimating worker mobility on the entire RTFL sample. This would entail two not negligible advantages: worker mobility measures would refer to the entire set of resident non-institutional employed; and they could be produced timely and updated quarterly.

However, the opportunity of using the RTFL for current estimation of worker mobility is conditional upon further checks. One has to ascertain if the relationships between mobility measures from the two sources found for 1995 extend to more recent years, characterized by increasing mobility. Research on this topic is underway.

References

- Baretta P., Trivellato U. (2003) *La mobilità dei lavoratori da fonti amministrative e da 'surveys' sulle famiglie: un'analisi comparata*, Working Paper 58, Progetto MIUR "Dinamiche e persistenze nel mercato del lavoro italiano ed effetti di politiche", Dipartimento di Scienze Statistiche, Università di Padova, Padova.
- Contini B. (Ed.) (2002) *Labour mobility and wage dynamics in Italy*, Rosenberg & Sellier, Torino.
- Davis S. J., Haltiwanger J. (1999) Gross Job Flows, in: *Handbook of Labor Economics*, vol. 3B, Ashenfelter O. & Card D. (Eds.), North-Holland, 2711-2805.