

IARIW 2010

“Parallel Session 2B: Surveys and Data Combination”

Title: Alternative measures of well-being: combining survey-based and administrative-based estimates.

Corresponding author: Lisa Crosato,

Statistics Department

Università di Milano-Bicocca

Via Bicocca degli Arcimboldi, 8, 20126 Milano

email: lisa.crosato@unimib.it

Co-authors: Paolo Mariani, Mauro Mussini and Biancamaria Zavanella

Abstract

Administrative data have been lately exploited not only for the study of income distribution (Atkinson, 2007), but also for data validation and evaluation of error measure in sample survey (Jäckle et al. 2004, Dragoset and Fields 2006).

In this paper we combine two data sources containing, among others, the main monetary proxies of well-being: income and consumption. Both sources, a sample survey and an administrative register, regard the municipality of Milan (Italy). Although international literature clearly points to this direction (Statistics Finland 2004, Jones & Elias 2006, Jenkins 2008), to our knowledge this is one of the first attempts of matching survey and administrative data in Italy (see ISTAT 2008).

The first dataset is the sample survey on family expenditure¹ conducted by the Milan Municipality and the Chamber of Commerce of Milan (wave 2005-2006). This survey is directed to a sample of about 800 families resident in the Milan Municipality and follows, both in sample selection and survey composition, ISTAT (Italian National Statistical Office) Household Expenditure Survey. The aim of the survey is therefore the estimation of the expenditures of Milan families in real goods and services. The second dataset collects administrative records regarding the entire population resident in Milan and covers a 7 year period, from 2000 to 2006. The data derive from tax records matched to the local population and family register in the data-warehouse AMeRiCA² (Mariani, Mezzanzanica e Zavanella, 2006 a and b), which contains more than 800,000 individual records per year. Each individual is characterized by yearly income as appearing in income tax return, divided according to income source (salary income, including pensions and subsidies, private income, estate income and income from other sources such as capital gains or rental income.) The integration with the population register represents one of the main advantages of AMeRiCA, since it allows reconstruction of families and implies the total coverage of the population either considering individuals or families. Furthermore, several socio-demographic characteristics are recorded for tax-payers and not. This makes it possible also a thorough characterization of people in principle exposed to poverty, each year and over time (Crosato e Zavanella, 2008; Minotti, Mussini e Zavanella, 2008).

This paper concerns the linkage between the above sample survey and the administrative register. We also try to give a few insights on the research issues raised by the new dataset characterizing families and individuals on both the expenditure and the earning side.

¹ “Indagine sui consumi delle famiglie nel comune di Milano”

² AMeRiCA (“Anagrafe Milanese e Redditi Individuali con Archivio”) is managed by the Statistics Department of Milan-Bicocca University on the behalf of Milan Municipality.

Firstly, we focus on data combination, referring to the literature which offers several tools (see Belin and Rubin, 1995, Gill, 2001, Ridder e Moffit, 2007, Jenkins et al. 2008). Exact or probabilistic matching is the most appropriate technique in our case, where not only a large part of statistical units belong to both datasets, but one (the sample survey) is included in the other. Since the original observation units are different in one dataset with respect to the other (families in the sample survey and individuals in the administrative register) we proceed through a first match of the survey households with “administrative” families based on address and number of components and through a second combination among individuals based on date of birth, sex, address and type of job. Disclosure and privacy issues affect only partially our research since data from the sample survey in our possess do not contain personal information such as fiscal code or family names and the administrative records are identified through a secreted code.

Secondly, with this paper we intend to contribute to the debate on definition and measurement of well-being offering a systematic comparison of different levels of well-being as measured by income (retrieved from tax records) and consumption (obtained from the sample survey). We think this is a very important matter in Italy considering that, although consumption is generally considered as a better indicator of well-being (Meyer e Sullivan, 2003, 2007, 2008), policy interventions on poverty are almost exclusively based on tax record income. This become even more important since combining the two data sources we can evaluate expenditures of non tax-payers.

Notwithstanding the fact that the information contained in both datasets is confined to a single Italian municipality, we think results may have broader significance, especially on the methodological side. It is worth noting that both the survey next wave (2007-2008) and the administrative records (2007 and 2008) will be soon available in order to prosecute this work and improve linkage rates. In particular, the secreted identifying code characterizing individuals in the administrative source allows for additional information about individuals and family history from 2000 to 2006. Furthermore, this dataset may become a fertile field for both the construction and the study of univariate and multivariate well-being indicators and for the assessment of expenditure survey reliability.

- Atkinson, A.B., 2007, Measuring Top Incomes: Methodological Issues, in A. Atkinson and T.Piketty (eds) Top Incomes over the Twentieth Century: A contrast Between European and English Speaking Countries, Oxford, Oxford University Press.
- Belin, TR e Rubin DB, 1995, A Method for Calibrating False-Match Rates in Record Linkage, JASA, 90.
- Copas, JB e Hilton, FJ, 1990, Record linkage: statistical models for matching computer records, J. of the Royal Statistical Society (A), pp.287-320.
- Crosato L., Zavanella B, 2008, L'evoluzione del reddito dei cittadini milanesi (2000-2004) sulla base di archivi amministrativi, Quaderni di Dipartimento di Statistica dell'Università di Milano Bicocca, n.24.
- Dragoset L.M. and Fields G.S., 2006, U.S. Earnings Mobility: Comparing Survey-Based and Administrative-Based Estimates, ECINEQ working paper 2006-55.
- Gill, L., 2001, Methods for automatic record matching and linking and their use in National Statistics. National Statistics Methodological Series No. 25. London: Office for National Statistics.
- ISTAT, 2008, Combining survey and administrative data in the Italian EU-SILC experience: positive and critical aspects, UNECE- Work session on statistical data editing, WP-14, retrieved in August 2009 from <http://www.unece.org/stats/documents/2008/04/sde/wp.14.e.pdf>.
- Jäckle A., Sala E. Jenkins S.P. e Lynn P., 2004, Validation of Survey Data on Income and Employment: the ISMIE Experience, ISER Working Papers, N. 2004-14.
- Jenkins SP, Lynn P, Jäckle A., Sala E, 2008, The Feasibility of Linking Household Survey and Administrative Record Data: New Evidence for Britain, Int. J. of Social Research Methodology, 11, pp 29-43.
- Jones, P. and Elias, P., 2006, Administrative data as a research resource: A selected audit. A report to the ESCR Research Resources Board.
Retrieved in August 2009 from

http://www.esrc.ac.uk/ESRCInfoCentre/Images/Administrative_Data_a_selected_audit_tcm6-25869.pdf

- Mariani, P., Mezzanzanica M. e Zavanella, B., 2006a, "Statistical Information Systems and Data Warehouses for Job Marketplaces" *Atti della XLIII Riunione Scientifica della Società Italiana di Statistica*, CLEUP, Torino 14-16 giugno 2006, pp.289-292;
- Mariani, P., Mezzanzanica, M. e Zavanella B., 2006b, Sistema informativo statistico per il supporto decisionale: il progetto AMERICA - Anagrafe Milanese E Redditi Individuali Con Archivio, in *Metodi Modelli e Tecnologie dell'informazione a supporto delle Decisioni*. A cura di P. Amenta, L. D'ambra, M. Squillante, e A. Ventre, Pubblicazioni DASES (Dipartimento di Analisi dei sistemi economici e sociali Università degli Studi del Sannio). Franco Angeli.
- Meyer BD e Sullivan JX, 2003, Measuring the well-being of the poor using income and consumption, *Journal of Human Resources* pp. 1180-1220.
- Ridder, G e Moffit, R, 2007, *Econometric Methods for Data Combination*, Handbook of Econometrics.