rivista di statistica ufficiale

n. 1
2012

Temi trattati

Istat
Istituto Nazionale
di Statistica

rivista di statistica ufficiale

n. 1
2012

Temi trattati

Istat

Per contattare la redazione o per inviare lavori scrivere a:
Segreteria del Comitato di redazione della Rivista di Statistica Ufficiale
All'attenzione di Gilda Sonetti
Istat – Via Cesare Balbo, 16 – 00184 Roma
e-mail: rivista@istat.it

rivista di statistica ufficiale

n. 1/2012

Periodico quadrimestrale
ISSN 1828-1982

Registrato presso il Tribunale di Roma
n. 339 del 19 luglio 2007

Istituto nazionale di statistica
Servizio Editoria
Via Cesare Balbo, 16 – Roma

Stampato nel mese di Ottobre 2012
presso il Centro stampa dell'Istat
Via Tuscolana 1788 – Roma
Copie 230

Regression-based approaches for the decomposition of income inequality in Italy, 1998-2008

Rosalba Manna[1], *Andrea Regoli*[2]

Abstract

Decompositions by population subgroups and by income sources represent the traditional techniques for decomposing income inequality. Compared with the classical methodologies, the regression-based method gives the opportunity of quantifying the contribution to the inequality of a set of factors, while taking the correlations among them into account. In this framework, two regression-based decomposition methodologies are used: the Fields method and the Shapley value approach, with the aim of measuring the relative contributions of individual as well as household factors to inequality in individual disposable incomes. The factors are introduced as explanatory variables in an income generating model that is estimated through a panel data regression model with time-invariant unobserved random effects. The results suggest that the most relevant factors in explaining the observed income inequality are gender, human capital as well as non-human capital whereas the work status and the area of residence only affect income differentials in a marginal way.

Keywords: Inequality decomposition; regression-based methods; Shapley value; panel data models.

1. Introduction

This work addresses a relevant and topical issue: inequality in the Italian income distribution. Recently, inequality has raised growing concerns at the global level as well as in the Italian society, where income differences are widening against the background of a deep macroeconomic recession and the entailed negative perception of the economic and financial situation at the household level. Moreover, the family background in Italy can powerfully limit the chances of moving up the social ladder, with the consequence that those who lack resources are likely to be disadvantaged in terms of opportunities too.

A number of recent studies have analyzed the evolution of household income inequality in Italy. Boeri and Brandolini (2004) found evidence of significant distributive changes across socio-economic groups despite inequality at the aggregate level was stable between 1993 and 2002. Through a decomposition by population subgroups defined by the

[1] Post Ph.D student (University of Naples "Parthenope", Department of Statistics and Mathematics for Economic Research), e-mail: rosalba.manna@uniparthenope.it.
[2] Associate Professor (University of Naples "Parthenope", Department of Statistics and Mathematics for Economic Research), e-mail: andrea.regoli@uniparthenope.it.

occupational status of the household head, the authors concluded that the income distribution shifted to the advantage of the self-employed and managers and to the disadvantage of the employees.

For the time period 1991-2004, Quintano *et al.* (2009) confirmed that, from 2000 on, a marked segmentation of households emerged, with widening gaps in the average incomes between the group of managers and self-employed and the group of employees. Moreover, a decomposition by income sources added new evidence upon the role played by the different sources in accounting for both the level of inequality and its trend. The increase in wage differentials was deemed to be the main driving force behind the dramatic rise in inequality in disposable household incomes between 1991 and 1993. The peak in inequality observed in 1998 was driven by the income from financial assets, whereas in more recent years the income from self-employment was found to be the main disequalizing factor.

The disappearance of the middle class was the main implication of Massari *et al.* (2009) work, which investigated the differences across the whole income scale by comparing the household income distributions in 2002 and in 2004 through a nonparametric approach, based on the relative distribution density function. The authors found an increased income polarization, due to downgrading of the incomes earned by the households headed by employees or by the self-employed.

Unlike the above mentioned studies, the present paper focuses upon the determinants of the observed income differentials. More precisely, the aim of this study is mainly empirical and it has to do with the assessment of the contribution of several individual and household factors to income inequality among individuals through a regression-based decomposition strategy.

A wide literature exists on the decomposition of inequality measures. The traditional methods include the decomposition by income sources (Shorrocks, 1982) and by population subgroups (Shorrocks, 1984). The former method estimates the contribution of individual income components to the observed inequality, whereas the latter allows to measure inequality both within and between subgroups of the population. Both of them are typically descriptive methods that tell us what sources of incomes or subgroups account for inequality but they fail to detect and measure the contributions of individual determinants to income inequality. For this reason, the information provided by those methods is of limited usefulness for policy-makers seeking to address income inequality problems.

Unlike the traditional methods, the regression-based approaches followed in this work have the advantage of going beyond decomposing inequality simply in terms of income components or discrete population categories. Indeed, they enable to include any factor that may drive the observed inequality, such as economic, social, demographic and policy variables, both discrete and continuous. Moreover the regression-based methods can manage problems of endogeneity due to reverse causality.

The regression-based decomposition methodology was proposed in the early 1970s (Blinder, 1973; Oaxaca, 1973), but failed to arouse much interest until Morduch and Sicular (2002) and Fields (2003) devised a regression-based decomposition by income determinants through the extension of the decomposition by income sources. Regression-based decompositions start with the estimation of an income-generating function, and then use the estimated coefficients to derive the inequality weight of every explanatory variable.

In the context of regression-based decomposition, many recent studies proposed the application of either the Fields method or the Shapley value approach, a concept taken from cooperative game theory.

Sastre and Trannoy (2002) measured the impact of different income sources on income inequality for UK and USA household income data, focussing on some methodological issues regarding the Shapley decomposition of Gini index.

The studies by Wan (2004), Wan and Zhou (2005) and Wan *et al.* (2007) combined the Shapley value approach and the regression-based decomposition technique in order to disentangle the contribution of different factors to household income inequality in China by using several inequality indices.

Israeli (2007) suggested a method for decomposing the R-Square of a linear regression that combines the Shapley approach with the Fields method and added an empirical illustration of this methodology on Israeli earnings data.

Guanatilaka and Chotikapanich (2009) investigated the evolution of Sri Lanka's expenditure inequality as well as its underlying causes by using three regression-based methodologies of decomposition: the Fields approach, the Shapley value decomposition and the Yun method.

Devicienti (2010) applied a Shapley value-based methodology for decomposing changes in the Italian wage distribution by using WHIP (Worker History Italian Panel) data on employees in private firms for the years between 1985 and 1999. The only other decomposition analysis that employed the Shapley value approach on Italian data is the recent study by Celidoni *et al.* (2011) who investigated the determinants of expenditure inequality on a pseudo panel based on Istat Household Budget Survey data for the years from 1997 to 2004.

In the wake of these studies, the present paper intends to contribute to the identification of the main driving factors for the inequality levels through the application of regression-based decomposition approaches. Unlike the above mentioned empirical studies for Italy, however, in this contribution the inequality is measured on individual incomes. Heterogeneity across individuals and across time is accounted for by using the longitudinal information from the Historical Archives of Bank of Italy's Survey of Household Income and Wealth.

The comparative discussion of the results derived from the application of Fields and Shapley approaches is also a key contribution of this paper.

This paper is organized as follows. In Section 2 the theoretical background of the regression-based methods is presented. Section 3 deals with model selection and specification issues, whereas the empirical data from the Survey of Household Income and Wealth (SHIW) are illustrated in Section 4. Section 5 shows the model estimates and the decomposition results, while conclusions are drawn in the final Section 6.

2. The regression-based decomposition according to the Fields method and the Shapley approach

Generally speaking, the regression-based inequality decomposition methods allow quantifying the impact of the determinants of inequality. Both the number and the kind of the explanatory factors are arbitrary, introducing some flexibility in the analysis that is not granted by the traditional decomposition methods.

Let us consider an income generating function such as:

$$\ln y = \sum_{j=1}^{k} b_j X_j + \varepsilon \tag{1}$$

where y denotes income, X_j the *j*-th explanatory variable, b_j its coefficient and ε the error term. The Fields method (Fields, 2003) estimates the share of the log-variance of income that is attributable to the *j*-th explanatory factor (the relative factor inequality weight) as:

$$s_{j,FIELDS} = \frac{\hat{b}_j \cdot \mathrm{cov}(X_j, \ln y)}{\sigma^2(\ln y)} \tag{2}$$

where $\hat{b}_j$ is the coefficient of the *j*-th explanatory factor estimated from an OLS multiple regression, $\sigma^2(\ln y)$ is the variance of the dependent variable and $\mathrm{cov}(X_j, \ln y)$ is the covariance between the *j*-th factor and the dependent variable.

The sign of s_j indicates whether the contribution of factor x_j is inequality-increasing $(s_j > 0)$ or decreasing $(s_j < 0)$. It holds that

$$\sum_{j=1}^{k} s_{j,FIELDS} = \frac{\sum_{j=1}^{k} \hat{b}_j \cdot \mathrm{cov}(X_j, \ln y)}{\sigma^2(\ln y)} = \frac{\sigma^2(\ln \hat{y})}{\sigma^2(\ln y)} = R^2 \tag{3}$$

When the error term ε of the regression is considered, its inequality contribution is given by the proportion of inequality unexplained by the explanatory variables included in the income regression, that is:

$$s_\varepsilon = 1 - R^2 \tag{4}$$

Under some assumptions, Fields extended this result to any inequality index with certain properties, including the most common measures such as the Gini index and the indexes belonging to the generalized entropy family. One limitation of the Fields method is that the functional form for the income generating function must be log-linear.

Unlike the Fields method, the Shapley value approach, as introduced by Shorrocks (1999), yields an exact additive decomposition of any inequality measure into its contributory factors. Indeed, the decomposition of a given inequality measure through a regression-based method combined with the Shapley value approach aims at assessing the contributions of a set of factors (the explanatory variables in the income regression model 1) whose sum accounts for the inequality indicator. Moreover, the income generating model can have any functional form (including linear, logarithmic and semi-logarithmic functions).

As in the framework of a general decomposition problem, the inequality measure calculated on the predicted income values $I(\hat{y} \mid X_1, X_2, ..., X_k)$ is expressed as the sum of the contributory factors:

$$I(\hat{y} \mid X_1, X_2, ..., X_k) = \Phi(X_1, I) + \Phi(X_2, I) + ... + \Phi(X_k, I) \quad (5).$$

The rationale behind the Shapley approach is that the contribution of a single factor can be assessed as the difference between the overall income inequality and the inequality that would be observed should that factor be removed from the set of income determinants. As a consequence, the marginal impact of each factor $\Phi(X_j, I)$ $j = 1,2,...,k$ is calculated through the estimation of a sequence of regression models starting from the specification which includes all the regressors, and then successively eliminating each of them. The overall marginal contribution of each variable is then obtained as the average of its marginal effects: since the contribution of any factor depends on the order in which the factors appear in the elimination sequence, this average is calculated over all the possible elimination sequences.

The contribution $\Phi(X_j, I)$ of the factor X_j to the explanation of the inequality measure I is given by the following formula:

$$\Phi\left(X_j, I\right) = \frac{1}{k!} \sum_{\pi \in \Pi_k} \left[I\left(\hat{y} \mid B\left(\pi, X_j\right) \cup \left\{X_j\right\}\right) - I\left(\hat{y} \mid B\left(\pi, X_j\right)\right)\right] \quad (6)$$

where $I(\hat{y} \mid X)$ is the inequality indicator calculated on the predicted income values from the regression on the vector of explanatory variables X ;
Π_k is the set of all the possible orderings (permutations) of the k variables;
$B(\pi, X_j)$ is the set of the variables preceding X_j in the given ordering π .

The calculation of each factor's contribution requires the estimation of $2^k - 1$ income generating models, and then the derivation of the inequality indicator I using the income predicted values for every model.

Finally, the proportion of unexplained inequality $I_R(y)$ is obtained as the difference between the inequality measure calculated on the observed income values $I(y)$ and the same measure calculated on the predicted income values, as follows:

$$I_R(y) = I(y) - I(\hat{y} \mid X_1, X_2, ..., X_K) \quad (7).$$

The relative inequality weight of the factor X_j may be written as:

$$s_{j,SHAPLEY} = \frac{\Phi(X_j, I)}{I(y)} \quad (8)$$

such that

$$\sum_{j=1}^{k} s_{j,SHAPLEY} = \frac{I(\hat{y} \mid X_1, X_2, ..., X_k)}{I(y)} \quad (9).$$

As pointed out by Israeli (2007), when income is expressed in log terms and the variance is used as inequality index, the Shapley decomposition according to formula (9) matches the Fields decomposition of the R-square according to formula (3) since

$$\sum_{j=1}^{k} s_{j,SHAPLEY} = \frac{I(\hat{y} \mid X_1, X_2, ..., X_k)}{I(y)} = \frac{\sigma^2(\ln \hat{y})}{\sigma^2(\ln y)} = R^2 \quad (10).$$

This does not mean that the factor contributions evaluated through the two approaches coincide. They do so only in the absence of correlation between the explanatory variables.

3. Model selection

The first step in the regression-based decomposition of income inequality requires the specification and the estimation of an income generating function, that is a model where income is regressed on some explanatory variables accounting for individual and household characteristics. For the estimation of the income generating function, we decided to exploit the potential of panel data by pooling the observations on a cross-section of individuals over several time periods.

We specified a panel data regression model with time-invariant unobserved effects (Wooldridge, 2002), which can be written as:

$$\ln y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it} \quad t = 1,2,...,T \quad i = 1,2,...,N \quad (11)$$

where $\mathbf{x}_{it}$ is a $1 \times K$ vector of regressors, c_i is the time-constant, individual-specific effect and u_{it} is the disturbance term for which the strict exogeneity condition is assumed to hold, that is

$$E(u_{it} \mid \mathbf{x}_{i1}, ..., \mathbf{x}_{iT}, c_i) = 0 \quad t = 1,2,...,T \quad (12).$$

This assumption implies that each error term u_i is uncorrelated with the regressors at all time periods, namely

$$E(\mathbf{x}'_{is} u_{it}) = 0 \quad s,t = 1,2,...,T \quad (13).$$

The two core specifications of such models are known as Random Effects (RE) and Fixed Effects (FE) models. In particular, we have specified a RE model, where the individual effect c_i is treated as a random variable that adds to the error term u_{it}. This choice is justified primarily by the RE model using both the “between variation” (the variability across individuals) and the “within variation” (the variability over time). For this reason, unlike the FE model, it allows both to estimate the coefficients of the regressors that do not vary at all over time (with null within variation) and to measure with no efficiency loss the effects of regressors that display a small within variation. In this study, the dependent variable is represented by the (log of) individual net disposable income, whereas the regressors include, among others, gender (that is invariant over time) and the years of completed studies (that exhibit a little variation over time).

Our preference for the RE model is also explained by the fact that we are not interested in estimating the values of the unobserved term for some specific individuals, but instead we concentrate on the influence of individual and household factors on the disposable income of hypothetical individuals with given characteristics. In the situation where the individuals are drawn randomly from a large population, as is usually the case for household panel studies, the RE model is an appropriate specification (Baltagi, 2008).

The RE estimator is derived under the further assumption of uncorrelation (orthogonality) between the individual effect c_i and the observed explanatory variables $\mathbf{x}_{it}$:

$$E(\mathbf{x}'_{it} c_i) = 0 \qquad t = 1,2,...T \tag{14}$$

This means that all the regressors $\mathbf{x}_{it}$ are considered to be exogenous. Many applications of the income generating function in a longitudinal framework have studied the effects of human capital accumulation on individual wages through the specification of a panel data model where the unobserved, individual term was intended to capture such features as individual ability. According to those studies, unobserved heterogeneity among individuals is likely to be correlated with some observed explanatory variables, which can generate potential endogeneity problems. Our study differs in that it is not restricted to the analysis of the returns to schooling and/or work experience on a sample of employees but it focuses on the estimation of the impact of several factors (that include both human and non-human capital as well as individual attributes) on the inequality in a measure of living conditions.

The above statement drove the choice of defining this measure as the individual disposable income, made up of wages, income from self-employment, transfers and income from capital. Traditionally the economic inequality is evaluated on household equivalent income or consumption, if one is willing to assume that the well-being of an individual depends on the combined resources of all the household members. The choice of measuring the inequality on individual income and, consequently, defining the individual as the unit of analysis is motivated here by the interest in explaining the determinants of inequality in the individual capacity to earn income, regardless of how the individual resources may be pooled together and then shared within the household.

4. Data, variables and summary statistics

The data used in this work are drawn from the Survey of Household Income and Wealth (SHIW) conducted every two years by the Bank of Italy on a sample of about 8,000 Italian households.

For every survey, the sample is composed of both households that have been already interviewed in previous years (panel households) and fresh households.

This survey is the only relevant source at the national level for household and individual longitudinal income data over a relatively large time interval. In particular, we referred to the Historical Database of the survey (Banca d'Italia, 2010) from which we selected information on the income earners born between 1938 and 1980 who have been successfully interviewed from 1998 to 2008 (they were between 18 and 60 years old in 1998). Such information took the form of a balanced micro panel where a large number of individuals N (N=1226) have been observed over a short time period T (T=6 years covering on the whole a time span of 10 years).

The selection of the individual longitudinal data was not an easy task. In the SHIW database every household is assigned a fixed identification number across waves whereas a fixed personal identification number for the individuals is missing. At every wave an individual is identified by both the household number and the order number of the individual within the household. The possibility of linking longitudinally the personal information exists however for couples of subsequent waves. Therefore the construction of the longitudinal dataset was achieved through the merging of individual data collected for couples of subsequent waves (1998-2000, 2000-2002, 2002-2004,...) followed by a further check for basic personal attributes such as gender and birth year.

In a longitudinal framework across waves, the situations of exits from the sample and/or later entries in the sample are not easily manageable since it cannot be ruled out that different individuals not constantly present in a household may be given the same order number within the household.

This remark drove the selection of a balanced panel (where only the income earners who participated continuously in the survey were included) though this choice may raise some theoretical issues relating to the attrition bias that could be addressed more deeply.

The choice of the semi-log functional form along with the selection of the explanatory variables were informed by the human capital theory suggesting that the ability to earn income is influenced by educational level and age. Gender is expected to play a special role in Italian income inequality due to the large gaps between men and women in the economic participation and opportunity, especially with reference to wage equality for similar work (Hausmann *et al.*, 2010). The remaining individual factors introduced as explanatory variables in the income generating model are work status (separating those who are employed and presumably receive an income from work from those who are not employed and whose income comes from other sources) and position in the household (accounting for whether or not the individual is the head of the household, according to his/her declaration). A measure of household wealth was also included accounting for the stock of non-human capital that is supposed to generate flows of income in the form of interests or rents. The net household's wealth is defined as the sum of real assets (property, businesses and valuables) and financial assets (deposits, securities, shares, etc) net of financial liabilities (such as mortgage loans and other debts). Then the geographical area of residence has been considered.

Table 1 - Descriptive statistics

VARIABLE	Definition	Obs	Mean	Std. dev.
logY	(Log of) Net disposable income	7356	9.709	0.763
Gender	=1 for male; =0 for female	7356	0.622	0.485
Education	Years of completed study	7356	10.259	3.840
Age	Age (in years)	7356	49.781	10.360
Household head	=1 for head of household =0 for other household member	7356	0.618	0.486
Work status	=1 for employed; =0 for not employed	7356	0.694	0.461
Geographical area	=1 for North and Centre; =0 for South and Islands	7356	0.725	0.4462
Household wealth	Real and financial wealth (in thousands of euro)	7356	268.330	369.917

Conditional on the information provided by the SHIW Historical Database, both the number and the kind of explanatory variables included as determinants seem to be broad enough to account for the main factors that are likely to explain income inequality.

Descriptive statistics for the variables introduced in the model are presented in Table 1.

The net disposable income is defined as the sum of individual income from wages, self-employment, pensions and other transfers, and property income, from both real and financial assets. Every income item is reported after tax and social security contributions. Negative or null income values were given null log (income) values.

5. Results

5.1 Model estimation

The Random Effects model in Table 2 shows the log of individual net disposable income as a function of demographic, human capital, work status, location and household wealth variables. The reported regression coefficients come from the estimation of the saturated model, that is the model including all the explanatory variables. Robust standard errors are computed in order to correct for potential heteroscedasticity.

Table 2 - Random effects model estimation

EXPLANATORY VARIABLE	Coefficient (robust std error)
Gender	0.3803*** (0.0312)
Education	0.0459*** (0.0032)
Age	0.0240*** (0.0018)
Head of household	0.3325*** (0.0228)
Work status	0.3743*** (0.0407)
Geographical area	0.2174*** (0.0305)
Household wealth	0.0004*** (0.0000)
Constant	7.0676*** (0.1318)

R^2 overall = 0.3534
N=1226; T=6
Wald chi-squared(7)=1323.9; *p-value*=0.00
***: significant at the 1% level

The signs of the estimated coefficients are in line with the theoretical expectations. Significant income gaps are due to gender, level of education, age, position in the household, work status and area of residence: ceteris paribus, on average the males, the more educated, the oldest, the heads of household, the employed and those who live in northern or central regions enjoy higher income levels. Larger income flows are also associated with larger stocks of wealth.

An overall R^2 equal to 0.35 indicates a satisfactory fit of the income regression model, when compared with other studies on the same phenomenon. We might have improved the fit by including interaction terms, but this would have created some problems in correctly assigning the resulting effect to the variables included in the interaction term.

5.2 Decomposition results through Fields method

Since the RE estimator is equivalent to an OLS estimator applied to conveniently transformed variables (Wooldridge, 2002), the results of the decomposition analysis according to the Fields method have been obtained from the OLS regression of the transformed variables: the new variables are obtained by removing from the original ones a fraction θ of their average over time, where θ is estimated as a function of the variances of both the error and the individual effects term.

The inequality weight of each factor (column 1 of table 3) was calculated through the formula (2) as a function of the corresponding OLS coefficient, the covariance between the log income and the factor, and the variance of log income. The inequality weights associated sum up to 36.3, which is the value of R^2 from the above regression. The remaining proportion (63.7) is attributed to the residual term, which means that a large portion of inequality is not explained by the variables included among the income determinants. Column 2 reports the percentage contributions of each factor to the explained inequality level. The most important variables in determining the explained income inequality are gender (21.3%) and household wealth (21.1%), followed by educational level (19.4%) and household head (16.5%). Smaller weights are attached to working status (9.4%), age (8.8%) and geographical area (a bare 3.5%).

Table 3 - Factor contributions to inequality using the Fields method

FACTOR X_j	Factor inequality weight s_jx100	Percentage contribution net of residual
Gender	7.7	21.3
Education	7.0	19.4
Age	3.2	8.8
Head of household	6.0	16.5
Work status	3.4	9.4
Geographical area	1.3	3.5
Household wealth	7.7	21.1
Residual	63.7	
Total	**100.0**	**100.00**

5.3 Decomposition results through Shapley approach

The results from the inequality decomposition using the Shapley value approach are reported in Table 4.

Since the decomposition results are influenced by the choice of the inequality index, the estimates are presented for four inequality measures: Gini index, Theil index, the mean logarithmic deviation, and the variance of logarithms.

The table shows the contributions to the income inequality in absolute terms (first column) and in percent of both the observed inequality (second column) and the explained inequality (third column).

When using either Gini or Theil index, the contributions of individual and household factors altogether account for more than 80% of the observed inequality. The explained inequality is smaller both for the mean log deviation (57.8%) and especially for the variance of logarithms (36%). In the latter case, as expected, the percentage of unexplained inequality is very similar to that resulting from the Fields method. Indeed, when applied to the variance of log income, the Shapley value approach is equivalent to the Fields decomposition of the R-square. The actual differences in the factor contributions are due to the presence of correlation among the regressors, which is not accounted for by the Fields method.

Table 4 - Factor contributions to inequality using the Shapley method

FACTOR X_i	Inequality measure					
	Gini			Theil		
	s_jx100	In % of (2)	In % of (1)	s_jx100	In % of (2)	In % of (1)
Gender	5.5	16.8	20.4	2.9	14.8	18.4
Education	4.1	12.5	15.3	1.8	9.1	11.3
Age	5.3	16.4	20.0	0.8	3.9	4.9
Head of household	4.2	12.7	15.5	2.2	11.1	13.8
Work status	1.4	4.3	5.2	0.9	4.7	5.8
Geographical area	1.3	3.9	4.8	0.4	1.8	2.2
Household wealth	5.0	15.4	18.8	6.9	35.2	43.6
(1) Total Explained Inequality	26.7	82.0	100.0	15.8	80.7	100.0
Unexplained Inequality	5.9	18.0		3.8	19.3	
(2) Observed Inequality	32.6	100.0		19.5	100.0	

FACTOR X_i	Inequality measure					
	Mean log dev			Var log		
	s_jx100	In % of (2)	In % of (1)	s_jx100	In % of (2)	In % of (1)
Gender	2.6	12.2	21.3	5.0	8.6	24.0
Education	1.7	7.8	13.6	3.1	5.4	15.0
Age	1.9	8.8	15.3	4.7	8.1	22.4
Head of household	1.8	8.5	14.8	3.3	5.7	15.9
Work status	0.7	3.2	5.5	1.2	2.1	5.9
Geographical area	0.4	1.9	3.3	0.8	1.4	3.9
Household wealth	3.2	15.0	26.2	2.7	4.6	12.8
(1) Total Explained Inequality	12.2	57.3	100.0	20.9	36.0	100.0
Unexplained Inequality	9.1	42.7		37.3	64.0	
(2) Observed Inequality	21.3	100.0		58.2	100.0	

For Gini index as well as for the variance of log income, the factors that explain the largest part of income inequality are gender and age, whereas for the indexes belonging to the class of entropy measures (that is Theil index and mean log deviation) the main determinants are household wealth and gender. By comparing the weight of human capital and non-human capital factors, the human capital variables (age and education, jointly considered) show the highest contribution to the explained inequality but for the Theil index, for whom the relative contribution of household wealth is especially large (43.6%).

While apparently different, individual and household factors are strictly intertwined. Indeed the human capital endowments are quite strongly correlated with both real and financial assets of the family of origin.

For all the inequality measures, the contribution of position in the household is estimated between 14% and 16%, whereas the remaining variables - occupational status and geographical area - are much less important as determinants of the inequality. This seems to suggest that, once human capital, gender, wealth and position in the household are taken into account, whether an individual is unemployed or not, and whether he or she lives in the North or in the South, have only a minor impact on income differentials.

6. Conclusions

Unlike the traditional inequality decomposition methods, the regression-based approaches allow to measure the inequality contribution of any explanatory factor. For this reason, regression-based methods are able to highlight what factors are most important in determining the observed income differentials.

However, there is a portion of income inequality that is not captured by the explanatory factors. Whenever the R-square of the income regression model is not very high, the Fields method is expected to leave a large share of inequality unexplained. On the other hand the performance of the Shapley approach is not directly linked to the fit of the regression model, being evaluated through the marginal impact of each factor, which differs depending on the choice of the inequality index. In our analysis, what constitutes the first result, the unexplained percentage of inequality is much lower when the Shapley approach is used and the Gini or Theil measure is calculated (respectively 18% and 19.3%) than when the Fields method is used (about 63%).

Our results on the drivers of income inequality shed light on the dominant role of gender, human capital, and wealth. Whatever the decomposition method and the inequality measure, the gender is found to play a key role as a determinant of income inequality, its contribution being estimated between 18.4% and 24%. This is likely to be a distinctive feature of Italy, as pointed out by many comparative studies. In Italy, women are known to find difficulties in combining work and family duties, and for this reason their participation in the labour force is low. On the other hand, women who have a job earn on average lower salaries and have usually fewer opportunities to reach leadership positions than men with comparable skills. Further applications of inequality decomposition methods for cross-country comparisons would be needed in order to support this evidence.

Along with gender, the endowments of human capital (education and experience) as well as physical capital (household assets) are found to be crucial determinants of income differentials, too. The role played by the household wealth stock is of primary importance when the Shapley approach is applied to the generalized class of entropy measures and to Theil index in particular. This remark highlights another advantage of the Shapley value approach over the Fields method: the former procedure allows to evaluate whether the marginal impact of each factor is equally important for every inequality measure or else to stress the different sensitivity of every inequality index to the underlying factors; on the contrary, through the latter method, the decomposition results are the same for a large number of inequality measures. A further advantage in applying the Shapley approach is the fact that, in the presence of many explanatory variables that may be correlated (and this is the case for our analysis) the Shapley approach accounts for the correlation among the determinants whereas the Fields method does not.

In order to complete the comment on the role of the different determinants, we found that work status and geographical area play minor roles in explaining inequality. The small weight attached to the area of residence may be surprising since this result seems to controvert the ingrained belief that the North-South divide is a major driver of economic inequality in Italy.

Finally, the regression-based decomposition methods employed in this paper may be further enhanced in order to provide policy insights. If among the explanatory factors some policy-relevant variables are introduced, e.g. labour market or redistributive intervention policies, the results of the decomposition may then be used in order to assess what decisions would be more effective in fighting the causes behind income inequality.

References

Baltagi B.H. (2008), *Econometric analysis of panel data*, Third edition, John Wiley & Sons Ltd, Chichester, England.

Banca d'Italia (2010), Historical Database of the Survey of Italian Household Budgets, 1977-2008, SHIW-HD, version 6.0, February 2010, On line at: http://www.bancaditalia.it/statistiche/indcamp/bilfait/docum/Shiw-Historical-Database.pdf

Blinder A.S. (1973), "Wage Discrimination: Reduced Form and Structural Estimates", *Journal of Human Resources*, 8, pp. 436-455.

Boeri T., Brandolini A. (2004), "The Age of Discontent: Italian Households at the Beginning of the Decade", *Giornale degli Economisti e Annali di Economia*, 63, pp. 449-487.

Celidoni M., Procidano I., Salmasi L. (2011), "Determinants of inequality in Italy: An approach based on the Shapley decomposition", *Review of Applied Socio-Economic Research*, vol.1, n.1, pp. 63-69.

Devicienti F. (2010), "Shapley-Value Decomposition of Changes in Wage Distribution: A note", *Journal of Economic Inequality*, 8 (1), pp.199-212.

Fields G. (2003), "Accounting for Income Inequality and Its Changes: A New Method with Application to the Distribution of Earnings in the United States", *Research in Labor Economics*, 22, pp. 1-38.

Guanatilaka R., Chotikapanich D. (2009), "Accounting for Sri Lanka's Expenditure Inequality 1980-2002: Regression-Based Decomposition Approaches", *Review of Income and Wealth*, 55 (4), pp. 882-906.

Hausmann R., Tyson L.D., Zahidi S. (2010), *The Global Gender Gap Report*, World Economic Forum, Geneva.

Israeli O. (2007), "A Shapley Based Decomposition of R-Squared of Linear Regression", *Journal of Economic Inequality*, 5 (2), pp.199-212.

Massari R., Pittau M.G., Zelli R. (2009), "A dwindling middle class? Italian evidence in the 2000s", *Journal of Economic Inequality*, 7 (4), pp. 333-350.

Morduch J., Sicular T. (2002), "Rethinking Inequality Decomposition, with Evidence from Rural China", *The Economic Journal*, 112, pp.93-106.

Oaxaca R. (1973), "Male-Female Wage Differences in Urban Labour Markets", *International Economic Review*, 14, pp.693-709.

Quintano C., Castellano R., Regoli A. (2009), "Evolution and Decomposition of Income Inequality in Italy, 1991-2004", *Statistical Methods and Applications*, 18 (3), pp. 419-443.

Sastre M., Trannoy A. (2002), "Shapley Inequality Decomposition by Factor Components: Some Methodological Issues", *Journal of Economics*, 9, pp. 51-89.

Shorrocks A. F. (1982), "Inequality Decomposition by Factor Components", *Econometrica*, 50, pp. 193-211.

Shorrocks A. F. (1984), "Inequality Decomposition by Population Subgroups", *Econometrica*, 52, pp. 1369-85.

Shorrocks A. F. (1999), *"*Decomposition Procedures for Distributional analysis: A Unified Framework Based on the Shapley Value*"*, mimeo, University of Essex.

Wan G.H. (2004), "Accounting for Income Inequality in Rural China: a Regression-Based Approach", *Journal of Comparative Economics*, 32, pp.348-363.

Wan G. , Lu M., Chen Z. (2007), "Globalization and Regional Income Inequality: Empirical Evidence from Within China", *Review of Income and Wealth*, Vol. 53, No. 1, pp.35-59.

Wan G., Zhou Z. (2005), "Income Inequality in Rural China: Regression-Based Decomposition Using Household Data", *Review of Development Economics*, 9 (1), pp.107-120.

Wooldridge J.M. (2002), *Econometric Analysis of cross section and panel data*, The MIT Press, Cambridge, Massachusetts, London, England.

Sample size for the estimate of consumer price sub-indices with alternative statistical designs [1]

Carlo De Gregorio [*]

Abstract

This paper analyses the sample sizes needed to estimate Laspeyres consumer price sub-indices under a combination of alternative sample designs, aggregation methods and temporal targets. In a simplified consumer market, the definition of the statistical target has been founded on the methodological framework adopted for the Harmonized Index of Consumer Prices. For a given precision level, sample size needs have been simulated under simple and stratified random designs with three distinct approaches to elementary aggregation, founded on Carli, Jevons and Dutot formulae. Alternative temporal targets are also examined: the single monthly target, the whole sets of monthly and quarterly indices, the annual average and the annual link. Empirical evidence is finally provided, based on the elaboration of survey microdata referred to elementary aggregates - such as air transport and package holidays - characterized by high volatility within and between months.

Keywords: chained index; aggregation; sample size; sampling variance; air transports; package holidays.

1. Introduction

In recent years there has been a growing concern for a more explicit use of the concepts and tools of statistical inference to produce estimates of consumer price indices (CPI) and, in particular, to define the targets of the estimates in the fashion typical of statistical survey methods. Although this issue has never been at the core of CPI literature, the pioneering works on this subject date back to Banerjee (1956) and Adelman (1958), while systematic research on the sampling variance of the Laspeyres CPI index has been developed since mid-eighties: see for example the session dedicated to this issue at the 1987 joint ISI-IASS

[1] I wish to thank Alexandre Makaronidis, former head of the HICP unit at EUROSTAT, Keith Hayes (head of HICP methodology) and Jan Walschots, as well as the colleagues of the National Statistical Institutes participating to the EUROSTAT Task force on HICP sampling. I am also grateful to my colleagues at ISTAT: Paola Anitori for her helpful comments, criticism and suggestions, and Alessandro Brunetti for further observations. I am as well grateful to the other colleagues of the former CPI unit, namely Carmina Munzi (this paper is dedicated to her memory), Patrizia Caredda, Stefania Fatello, Rosanna Lo Conte, Maurizio Massaroni, Stefano Mosca, Francesca Rossetti and Paola Zavagnini. I finally acknowledge the useful comments and advices of an anonymous referee. Nevertheless, the views expressed here are solely those of the author and do not necessarily reflect those of ISTAT nor EUROSTAT. An earlier draft of this paper has been presented at the 12th meeting of the Ottawa group (International working group on price indices), held in Wellington (NZ) the 4-6 May 2011.

[*] Senior researcher (Istat), email: cadegreg@istat.it

conference and in particular the works of Biggeri et al. (1987), Andersson et al. (1987) and Leaver et al. (1987), or the works by Kott (1984) and more recently by Dalèn (2001).[2] Further research focused also on specific issues, such as the properties of sample designs based on alternative formulas for elementary aggregation (Dalèn 1992; Baskin et al. 1996; Fenwick 1998; ILO 2004, chap.5; Balk 2008, chap. 5). Stimuli for the adoption of more developed statistical techniques also came along with the innovations in price collection, especially in selected consumer markets: this happened with scanner data (De Haan et al. 1997; Fenwick 2001; Koskimäki et al. 2003) and with sources like e-commerce and administrative or private databases. In general, the availability of larger and more flexible data sets of price quotes made it necessary to set up generalized methods, fostering a greater attention on sampling issues. The integration with other statistical sources also favoured the adoption of sampling based approaches: for instance, the availability of regularly updated business registers has been considered to improve sample design in Biggeri et al. (2006).

Nevertheless, most of the empirical approaches adopted to measure the variance of the estimates relied on the use of replication techniques, since the data available for analysis derive mainly from purposive samples and quite rarely from probabilistic designs. Several authors dealt with this issue (Biggeri et al. 1987; Andersson et al. 1987; Balk 2008, p.176) and with the need to provide a suitable statistical design (Kott 1984; Dalèn et al. 1995; Dalèn 1998, 2001; Ribe 2000; Dorfman et al. 2006). The definition of both the universe and the target parameters appears by far the most critical issue in a CPI sampling design and, more in general, in the CPI itself (Dalèn 1998, 2001). Several factors connected to the rapid evolution of consumer markets impair this definition: they are related to products and outlets replacements as well as to the changes in their characteristics, and they should be tackled, at least theoretically, in order to provide a solid foundation for the production of the estimates. Therefore, the need of a structured framework of concepts and definitions has emerged in order to reduce complexity, and to provide sufficiently general and operative solutions.[3]

Ribe (2000) and the most recent methodological developments of the chained Laspeyres index adopted in the EU Harmonized Index of Consumer Prices (hereafter HICP) propose to structure the universe of transactions into homogeneous partitions based on the concepts of product-offer and consumption segment. Starting from this approach, this paper proposes a definition of the target universe under some assumptions on the functioning of consumer markets (they help to simplify the identification of the statistical target), and provides a tool to measure the sample size needed to estimate the sub-indices of HICP under alternative designs. The results obtained with simple and stratified random designs are compared, taking into account the use of alternative criteria for segmentation, elementary aggregation and temporal targets. Two case studies are also developed.

Based on the annually chained Laspeyres formula used in the HICP, a definition of the statistical target for a monthly index is firstly provided (par. 2). Given the desired precision

---

2 The literature on this subject is briefly surveyed in Dalèn et al. (1995). See also Wilkerson (1967), Dippo et al. (1983), Leaver et al. (1987), Leaver et al. (1991), Baskin et al. (1996), Norberg (2004). For an overview of variance estimation approaches in selected countries see ILO (2004, chap.5).

3 "*The consumer market ultimately consists of an enormous (but finite!) number of transactions, where goods and services (products) are purchased by consumers. However, it is not feasible to compare transactions directly between periods. Like the physicists who divide matter successively into molecules, atoms and nucleons, we have to bring some structure into our market universe as a prerequisite for a measurement procedure.*" Dalèn (2001, p.3).

level, we derive the sample size with simple and stratified random designs under Carli, Jevons and Dutot aggregation. For each design, alternative temporal targets are examined such as monthly, quarterly and annual indices, as well as the annual link of the chaining sequence (par. 3). The approaches are then tested on an experimental ground, simulating artificial populations from the microdata relating to two sub-indices of the HICP - air transports and package holidays - both characterized by a high volatility of price dynamics within and between months (par. 4).

2. The statistical target

2.1. Some aspects of the construction of the HICP

The HICP is a monthly Laspeyres index based on the average of a reference year (*yr*).[4] It is built as a chained index by linking together the monthly price indices *I* of the current year *y* based on the link month of *December y-1* and the fixed base index *H* of *December y-1*. By iteration, in the reporting month *m* of year *y* the aggregate HICP is derived as the product of three elements: a fixed base index (*H*), the product of the *(y-yr-1)* annual links, and the link index of the reporting month (*m*). In formulas:[5]

$$H^{y,m} = H^{y-1,12} I^{y,m} = H^{yr,12}\left(\prod_{x=yr+1}^{y-1} I^{x,12}\right) I^{y,m} \quad (1).$$

The link of the reporting month $I^{y,m}$ is compiled as the weighted average of the sub-indices referred to an exhaustive set of disjoint aggregates *j* of the total consumer expenditure in the weight reference year:

$$I^{y,m} = \sum_j I_j^{y,m} w_j^y \quad (2)$$

where the expenditure weights *w* add up to unity and in principle change every year since they are referred to the consumption expenditure of year *y-1*.[6]

The construction of any HICP aggregate follows a hierarchical procedure: the aggregation of the link indices comes first, then the result is chained to the fixed base index of the same aggregate. The sub-indices $I_j^{y,m}$ can be therefore interpreted as the primary components of the HICP and, as a consequence, each of them represents a distinct statistical target (Ribe 2000, p.1): hereafter we shall refer to the problem of estimating these sub-indices. Notice also that expression (2) can be applied to any exhaustive partition of the target consumption expenditure: we choose in particular to deal with the partition realized through the groups of COICOP-HICP classification at the lowest level of detail used for

[4] At present yr=2005.

[5] See EUROSTAT (2001, p.175-197). To simplify notation, in what follows the basis of all indices has been set to 1 instead of the usual 100.

[6] Furthermore, the weights are price-updated from the weight reference period (y-1) to the price reference period (December y-1). See EUROSTAT (2001, p.188-190), Hansen (2006), ILO (2004, chap.9).

HICP dissemination. This partition concerns almost 100 sub-indices (EUROSTAT 2001, p.253-68).

2.2. Re-pricing of transactions and statistical target in a HICP perspective

One of the main advancements in HICP methodology and legal basis regards the statistical definition of the HICP universe.[7] In particular, the target parameter for the annual links of a monthly Laspeyres CPI corresponds to the ratio of two simulated consumption expenditures obtained by mapping the universe of transactions in the weight reference period (year *y-1*) onto the sets of available offers in the price reference period (*December y-1*) and in the reporting period (month *m* of year *y*). In order to define this *re-pricing* of transactions, the concept of product-offer was introduced in EC Regulation 1334/2007: "*product-offer means a specified good or service that is offered for purchase at a stated price, in a specific outlet or by a specific provider, under specific terms of supply, and thus defines a unique entity at any one time"*. As a matter of fact, product-offers are the observation units in CPI sampling and they determine the partition of total transactions. Nevertheless they represent a rapidly changing stock: they may change as the characteristics of the goods and services evolve, as they are replaced, as retail evolves, or simply as prices change. In order to provide stable entities on which price comparisons can be based, the sets of all the transactions and product-offers in the statistical universe are exhaustively clustered into consumption segments, where each segment identifies homogeneous product-offers with regard to marketing targets, consumption purposes and characteristics. In the HICP framework, consumption segments represent the fixed objects to be followed by the Laspeyres index.[8].

The mechanism of re-pricing can be summarised on the basis of Figure 1, in which full information is assumed on consumer expenditure. The squared area on the left summarises the total consumer expenditure in the weight reference year *y-1* within a given consumption segment: the geometric shapes (circles, lozenges and hexagons) identify the product-offers, while the black smaller circles represent the actual transactions. Only a subset of the product-offers and transactions of year *y-1* is in common with the price reference month (that is *December y-1*) – the upper central side of Figure 1. Given this information, the consumer expenditure in year *y-1* is simulated by means of mapping functions connecting the product-offers available in that year with those available in the price reference month.[9] An identical approach is adopted for the re-pricing based on the reporting month (*y, m*): in this case there is no overlapping with transactions and product-offers in year *y-1*. The index

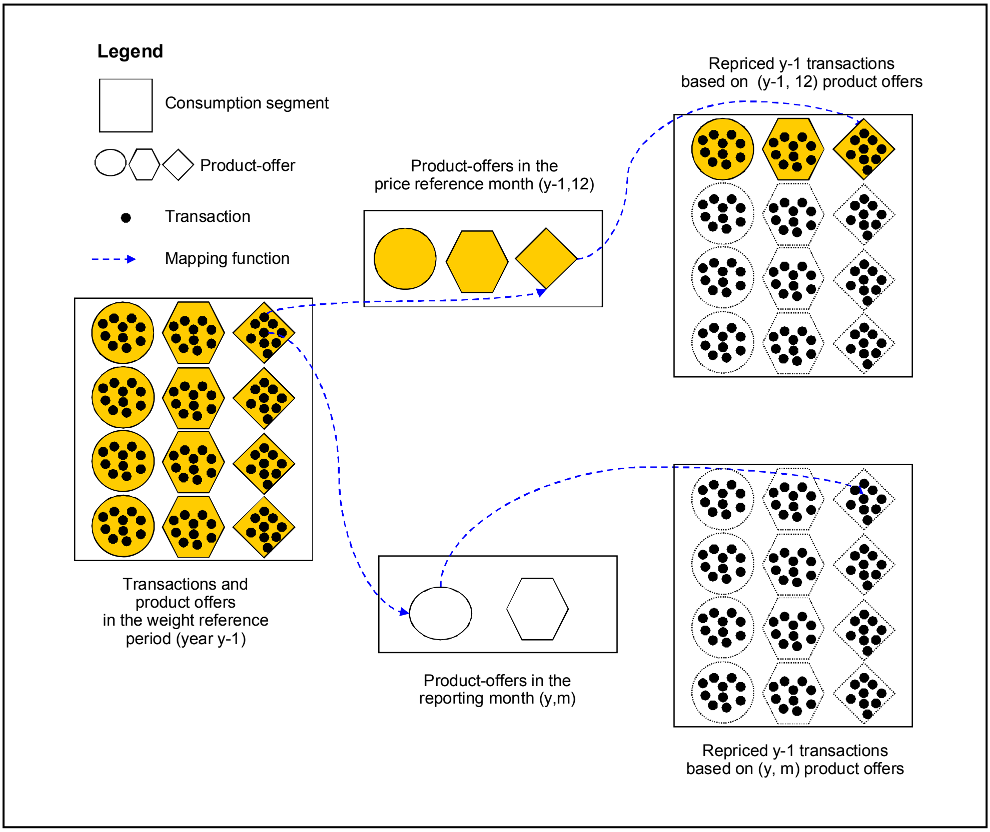
[7] See both Ribe, 2000 and Commission Regulation (EC) no. 1334/2007; a collection of the early HICP legislation can be found in EUROSTAT (2001).

[8] In particular: "*(...) 'consumption segment' means a set of transactions relating to product-offers which, on the grounds of common properties, are deemed to serve a common purpose, in the sense that they: are marketed for predominant use in similar situations, can largely be described by a common specification, and may be considered by consumers as equivalent. (….). The notion of consumption segments by purpose is therefore central to sampling and to the meaning of quality change and quality adjustment. However, an ambiguity in this concept concerns the level of aggregation at which it is defined and applied. (...) The range of product-offers will change over time as products are modified or replaced by retailers and manufacturers. The HICP requires the representation of all currently available product-offers within the consumption segments by purpose selected in the reference period in order to measure their impact on inflation. This applies particularly to new models or varieties of previously existing products.*" (EC Regulation 1334/2007).

[9] The common set of *December y-1* transactions is simply replicated (see the continuous border forms in the first row of the upper right square in Figure 1) while the rest of *y-1* expenditure is simulated (dotted border).

for the consumption segments is finally obtained as the ratio of the value of the two sets of simulated transactions.

Figure 1 - The re-pricing of transactions



Re-pricing has the advantage of providing a general framework for the provision of suitable solutions for the statistical treatment of inflation estimates. Mapping functions implicitly or explicitly incorporate various aspects of consumer behaviour modelling, and consequently define statistical imputation techniques, non response treatment, quality adjustment. They represent the methodological core of the estimates: their complexity directly depends on the rapidity of the changes that occur in the set of product-offers both generated by changes in the price level and by the range of the goods supplied to consumers. It is easily understandable that mapping functions are open to host several alternative hypotheses. As a matter of fact, the whole framework for the definition of HICP statistical universe is a theoretical tool open to a wide range of possible solutions, while methodological and empirical research is still needed in order to test its applicability as a statistical tool. Another key point is given by the definition of consumption segments. HICP regulation itself recognizes that ambiguities still concern the level of aggregation with which consumption segments are defined and applied. It is likely that consumption segments need to be specified case-by-case and this fosters the strategic role of consumer markets analysis, such as for example the structure of supply and demand, the marketing approaches and the segmentations adopted by producers and dealers. This part of the job strictly interacts with the definition of mapping functions.

2.3. The definition of the statistical target

Given the Laspeyres formula, and following the approach set up in Ribe (2000) and in HICP legal basis, the starting point for the definition of each target sub-index is given by the set of all the transactions in the weight reference year *y-1* concerning the COICOP-HICP group *j*. We assume a perfect knowledge of all information necessary to compile the indices. In particular, each transaction in the weight reference period is tracked; it concerns the purchase of a product-offer, and each product-offer is attributed to a specified consumption segment.

Product-offers are defined by the combination of two sets of characteristics. A first set consists of a vector g_i of variables describing the product, the outlet and the corresponding consumption segment (*h*): as a shortcut, we shall refer to such a vector with the term "product". Naming with $G = \{g_i \mid i = 1,...,N\}$ the set of all available products for the consumption purpose *j*, it is exhaustively divided in *M* disjoint consumption segments G_h, with *h=1,...,M*. This partition (Γ_M) can be expressed as follows (to economize notation, hereafter we omit the suffix *j*):

$$\Gamma_M = \{G_h, h = 1,...,M\},$$

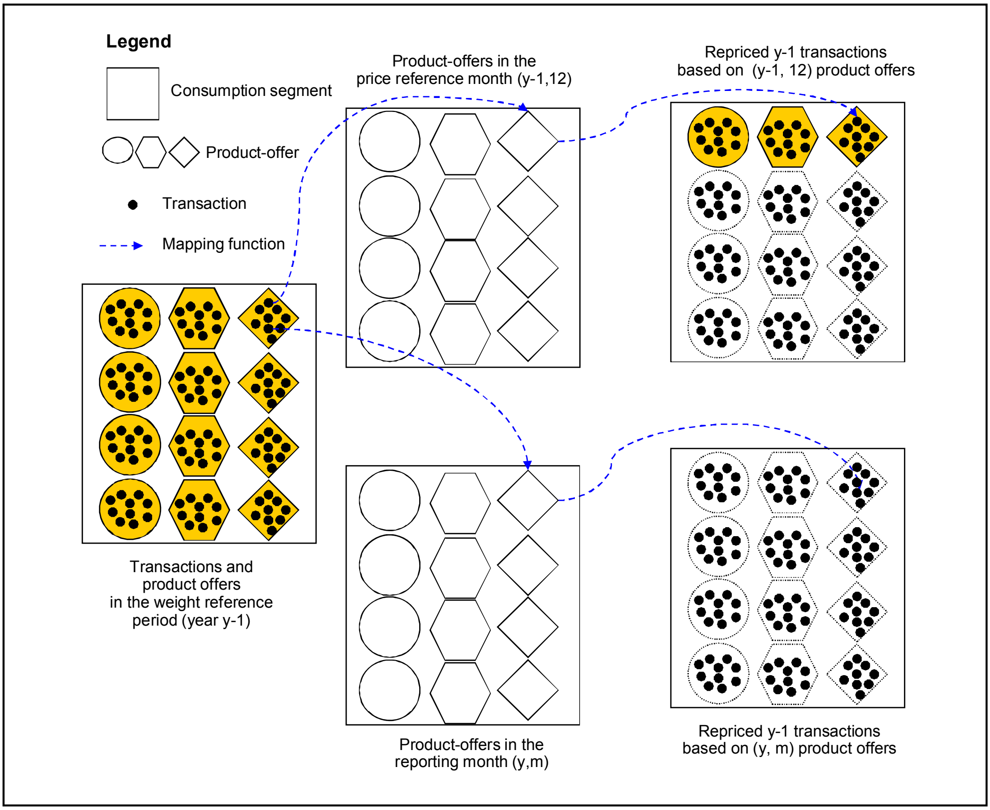
where $\bigcup_h G_h = G$ and $G_{h'} \bigcap G_{h''} = 0$ for every $h' \neq h''$. In order to simplify the definition of the universe and of the statistical target, some assumptions on the available product-offers are here introduced: the objective is to limit to price changes the possible sources of changes in the reference universe, and to provide a simplified framework for the definition of the statistical target (Dalèn 2001; Balk 2008, chap.5). The elements of the set *G* are assumed to be fixed and time-invariant: in other words, the number and the characteristics of the available offers do not change, and outlets and providers remain also unchanged (hypotheses A).

In this static environment, each vector g_i in the set *G* is associated to a second set of characteristics which describes the sequences of price spells and the corresponding time intervals of validity (p_i^t). From the definition of product-offer - recalled in par. 2.2 – each combination (g_i, p_i^t) describes a single product-offer. In order to control the number of product-offers associated to each product g_i we assume that discrete monthly pricing policies are adopted, where the prices of each element of *G* are eventually changed only at the very beginning of each month (hypotheses B).[10]

[10] This hypothesis may appear quite restrictive and not realistic, since monthly policies are quite rare. Nevertheless it is needed here in order to simplify notation and formalisation: the results can anyway be easily generalised, at least conceptually, to take account of intra-month policies.

Figure 2 describes the simplified framework derived from hypotheses A and B: it should be compared with the more general case reported in Figure 1. Product-offers can now be easily mapped given the invariance of G and the regularity of the sequence of price changes.

Figure 2 - A simplified framework for re-pricing



Given the hypotheses A and B, the generic element of the *(Nx13)* matrix Ω^{y-1} of all the product-offers in available year *y-1* is given by:

$$\left[g_i; \quad p_i^{y-1,1} \quad \dots \quad p_i^{y-1,12}\right].$$

The consumption expenditure (E) in the weight reference year *y-1* can be expressed as follows:

$$E^{y-1}\left(\Omega^{y-1}\right) = \sum_h \overline{p}_h^{y-1} T_h^{y-1} \tag{3}$$

where T labels the number of transactions and $\overline{p}_h^{y-1} = \sum_m p_h^{y-1,m}\left(\frac{T_h^{y-1,m}}{T_h^{y-1}}\right)$ is the annual average price actually paid for transactions T in the consumption segment h.

In order to define the "true" value of the target sub-index $I_j^{y,m}$ by adopting the consumption segments as the fixed objects defined in the HICP frame,[11] we need to simulate by means of re-pricing the total consumer expenditure of the weight reference year (*y-1*) on the basis of the product-offers available in the reporting month *m* of year *y* (identified by the couple (*y, m*)) and in the price reference month (*December y-1,* conventionally labelled with (*y,0*)). By applying (3) we obtain:

$$I^{y,m} = \frac{E^{y-1}\left(\Omega^{y,m}\right)}{E^{y-1}\left(\Omega^{y,0}\right)} = \frac{\sum_h \overline{p}_h^{y,m} T_h^{y-1}}{\sum_h \overline{p}_h^{y,0} T_h^{y-1}} = \sum_h I_h^{y,m} w_h \tag{4}$$

where $w_h^y = \frac{\overline{p}_h^{y,0} T_h^{y-1}}{\sum_h \overline{p}_h^{y,0} T_h^{y-1}}$ is the normalized value weight of consumption segment *h* and

$I_h^{y,m} = \frac{\overline{p}_h^{y,m}}{\overline{p}_h^{y,0}}$ is its price relative.[12] Average prices are derived on the basis of a mapping of the set of product-offers Ω^{y-1} into the sets $\Omega^{y,0}$ and $\Omega^{y,m}$ available in the price reference and reporting months. The hypotheses A and B relating to set *G* make it possible to assume the existence of a one-to-one correspondence through mapping functions connecting product-offers: for each transaction involving product g_i in year *y-1*, the corresponding product-offers in the base and reference years are $(g_i, p_i^{y,0})$ and $(g_i, p_i^{y,m})$ respectively. Different versions of the target parameter defined in (4) can now be provided adopting alternative aggregation methods to calculate average prices. Two alternative approaches are proposed here, namely the weighted arithmetic mean:

$$\overline{p}_h^{y,m} = \frac{\sum_{g_i \in G_h} p_i^{y,m} T_i^{y-1,m}}{\sum_{g_i \in G_h} T_i^{y-1,m}} \tag{5}$$

and the geometric mean:

$$\overline{p}_h^{y,m} = \exp\left[\frac{\sum_{g_i \in G_h} T_i^{y-1,m} \ln\left(p_i^{y,m}\right)}{\sum_{g_i \in G_h} T_i^{y-1,m}}\right] \tag{6}$$

[11] See EC Regulation 1334/2007.

[12] Non zero average prices by segment in the base month (y,0) are here assumed; on the treatment of zero prices in the HICP see EUROSTAT (2001, p. 184-5).

Each approach implies specific assumptions on consumers' elasticity to price changes (see below, section 3.1). The aim, then, is to compare the properties of alternative standard approaches to sampling in order to provide an estimate for $I^{y,m}$.[13]

3. Sample size with alternative designs and aggregation formulas

3.1. Simple random sampling (SRS)

Assume that a simple random sample *S* of *n* products is drawn from *G* and to collect the prices of the corresponding product-offers in the price reference month and in a generic reporting month *m*. Given the hypotheses A and B, this is equivalent to drawing an identical sample of products from the sets of available product-offers $\Omega^{y,0}$ and $\Omega^{y,m}$. No other information is available on the universe of product-offers, consumption segments and transactions.

Different estimates of I^m can be produced[14] depending on the approach followed to aggregate the sampled quotes and to produce the target index estimates. Three alternative types of frequently used unweighted means are here compared: Carli (arithmetic mean of price relatives, labelled with "C"), Dutot (ratio of mean prices, "D") and Jevons (geometric mean, "J"), respectively:

$$\hat{I}^{m,C} = \frac{\sum_{i \in S} \frac{p_i^m}{p_i^0}}{n} = \frac{\sum_{i \in S} I_i^m}{n} \tag{7}$$

$$\hat{I}^{m,D} = \frac{\sum_{i \in S} p_i^m \Big/ n}{\sum_{i \in S} p_i^0 \Big/ n} = \frac{\sum_{i \in S} I_i^m p_i^0}{\sum_{i \in S} p_i^0} \tag{8}$$

$$\hat{I}^{m,J} = \exp\left[\frac{\sum_{i \in S} \ln\left(I_i^m\right)}{n}\right] \tag{9}$$

The relative convenience of aggregation formulas has been deeply debated in literature and it has been evaluated on the basis of their economic properties and on the characteristics of the underlying distributions of price changes. Each approach entails in fact specific assumptions on consumers' elasticity to price changes, which are reflected on

[13] This methodological framework is clearly open to a larger set of different approaches to aggregation. Even the stochastic approach can be considered, although it has been largely criticised, mainly for its weak economic foundations. Particular conditions concerning the distribution of price changes might anyway spur the adoption of this approach.

[14] Hereafter, we drop the suffix labelling the year.

the implicit weighting of transactions (Ilo 2004, chap.9; Leifer 2002, 2008; Viglino 2003; Balk 2003, 2008, chap. 5; Silver et al. 2006): Carli aggregation implies equal value weights for the product-offers and, for each product-offer, a constant expenditure in both the price reference and in the reporting month.[15] The Dutot formula implies equal and time-invariant quantities for each product-offer, whilst the Jevons formula assumes that the expenditure shares of the price reference month do not change when relative prices change so that some substitution due to the change in relative prices is therefore implied. With respect to the Carli formula, the Dutot approach assigns a higher weight to the product-offers with a higher price level in the price reference month and the Jevons approach assigns a higher weight to the product-offers with a lower price dynamics. Without going into the issue of the choice of the "right" formula, we want to discuss here some of their statistical properties in terms of precision within different sampling designs.[16]

Formula (7) provides an unbiased estimator of (4)-(5) only if the probability of selection is proportional to the weight of each product (Adelman 1958). The same applies to (9) with respect to the target set by expressions (4) and (6). For the Dutot formula (8) to be unbiased with respect to (4)-(5) it is necessary to add the condition that the price relatives be independent from the price levels in the price reference month (Balk 2008, chap.5).

Given a confidence level α and a relative error expressed as a share of the sample mean ($r = \beta I^{m,q}$, where $q \in \{C, D, J\}$), the adoption of these three methods implies some differences in the sample sizes needed to produce an error lower than the $\beta\%$ of the true value of the parameter with a probability of $\alpha\%$. These differences depend on the standard errors of the three types of sample means and on the form of their distributions. By adopting standard simple random sampling theory (Cochran 1977, chap. 4-6) separately for the three aggregation formulas, the necessary sample size in month *m* in the case of the Carli formula can be expressed as follows:

$$n_{SRS}^{m,C}(\alpha,\beta) \geq \left[\frac{t_\alpha}{\beta} \frac{s_{\hat{I}}^{m,C}}{\hat{I}^{m,C}} \right]^2 = \left[\frac{t_\alpha}{\beta} \right]^2 C_{I^{m,C}}^2 \tag{10}$$

where t is the corresponding value of the t-Student distribution, $s_{\hat{I}}^{m,C}$ is the standard error of the sample Carli mean and $C_{I^{m,C}}$ is the coefficient of variation of the Carli index (Cochran 1977, sect. 4.6; Ilo 2004, chap.5).[17]

For the Dutot formula we obtain:

$$n_{SRS}^{m,D}(\alpha,\beta) \geq \left[\frac{t_\alpha}{\beta} \right]^2 \left[C_{p^m}^2 + C_{p^0}^2 - 2\rho_{0,m} C_{p^0} C_{p^m} \right] \tag{11}$$

[15] Notice that the systematic use of the Carli formula has been banned for the estimates of the HICP (Commission Regulation (EC) No 1749/96, Art.7; EUROSTAT (2001, p.129, 155-156)).

[16] For discussions on this issue see for example Fenwick (2008) and Baskin et al. (1996).

[17] Hereafter, the sample fraction correction is not considered.

where $\rho_{0,m}$ is the Pearson's correlation coefficient of the price levels in the price reference and in the comparison month, while C_{p^m} and C_{p^0} are the coefficients of variation of the price series in the two months (Cochran 1977, sect. 6.3-6.5; Ilo 2004, chap.5).

In the case of the Jevons aggregation, the following expression is derived:

$$n_{SRS}^{m,J}(\alpha,\beta) \geq \left[\frac{t_\alpha}{2\log(x_\beta)}\right]^2 \left(s_{\log(I)}^m\right)^2 = \left[\frac{t_\alpha}{2\log(x_\beta)}\right]^2 C_{I^{m,J}}^2 \tag{12}$$

where $x_\beta = \frac{+\beta + \sqrt{\beta^2+4}}{2}$, $s_{\log(I)}^m$ is the standard deviation of the logarithms of the individual indices I_i and it is equal to the coefficient of variation of the Jevons mean $C_{I^{m,J}}$ (Cochran 1977, sect. 4.6; Ilo 2004, chap.5). The expression for x_β can be derived by applying the SRS formula for confidence interval to the log transformed variable and then transforming back and resolving by n. Following Norris (1940), $\hat{I}^{m,J} s_{\log(I)}^m$ corresponds to an estimate of the standard deviation of the geometric mean.

Expressions (10)-(12) derive sample size from the product between two elements: one dependent on α and β, and the other one is based on the coefficients of variation of indices and – in the case of Dutot aggregation - price levels. For reasonably low values of β (e.g., lower than 10%), the comparison among these formulas can be limited to this last element. In general, when the variability of prices and indices is very small, the three approaches lead to very similar sample sizes. On the contrary, some important differences might emerge when the variability indices and price levels is relatively large.

The Dutot index needs a higher sample size when there is a strong heterogeneity in price levels with negative or low positive correlation between price levels, and in particular when the largest price changes are associated with goods with a higher price level in the price reference month. In the case of the Jevons formula, the sample size tends to be relatively higher if the distribution of the price changes is negatively skewed while the opposite happens with a positive skewness.

3.2. Stratified random sampling (STRS)

It is reasonable to expect that a partition in consumption segments can potentially isolate homogeneous product-offers and, once adopted as a stratification criterion, may consequently reduce differences in the aggregation formulas. Nevertheless, in principle partitioning in consumption segments may not represent a good stratification criterion, and in any case this may not be the best way to control the variability of price changes. These two concepts are clearly conceptually distinct but may nearly coincide if the "economic" criteria adopted to define consumption segments meet also, as a by-product, the objective of isolating clusters of products characterised by homogeneous pricing policies. Consumption segments should be based - according to the HICP legal based recalled in section 2.2 – on supply and demand side market analysis, and it is very likely then that they can target well the variability of price changes. The definition of the border between segmentation and stratification very much depends on the ambiguity – reminded also in HICP sampling regulation - on the level of aggregation, that is on how deep the segmentation is run. Such an issue deserves deeper case studies necessarily based also on survey microdata. Here we implicitly chose to collapse the two concepts and to compare the performance of different degrees of aggregation, from no stratification (segmentation) at all to deeper stratification.

We assume that more information is available concerning the consumption expenditure in the weight reference year: the true weighting structure (w_h) of a partition in consumption segments (Γ_M) is known, although no other information is available within each segment concerning the expenditure shares of the product-offers. It is important to notice here that weights are not identified here as a potential source of errors: this practice is common to most of the approaches to the measurement of the statistical error in CPI estimates (Biggeri et al. 1987 is a meaningful exception). In this work we adopt this same hypothesis, although we are perfectly conscious that additional work needs to be done in order to join this analysis of price and price indices variability with that of the precision of weighting: the latter is of paramount importance in order to evaluate the effectiveness of deeper stratifications.

If a stratified random design is adopted, the estimate may be obtained as a value-weighted arithmetic mean of the indices of each segment (stratum):

$$\hat{I}^{m,q}_{STRS} = \sum \hat{I}^{m,q}_h w_h \ .$$

The standard deviations within each stratum, for the three alternative formulas, will be given by:

$$s_h^{m,C} = \frac{\sum_{i \in h}\left(I_i - \hat{I}_h^{m,C}\right)^2}{n_h}$$

$$s_h^{m,D} = \frac{1}{n_h\left(\bar{p}_h^0\right)^2}\left[s^2_{p_h^m} + \left(\hat{I}_h^{m,D} s_{p_h^0}\right)^2 - 2\hat{I}_h^{m,D} s_{p_h^0 p_h^m}\right]$$

$$s_h^{m,J} = \hat{I}_h^{m,J} s^m_{\log(I_h)}$$

Independently of the type of elementary aggregation, total sample size with optimal allocation can be expressed as follows:

$$n^{m,q}_{STRS}(\alpha,\beta) \geq \left[\frac{t_\alpha}{\beta}\right]^2 \left[\frac{\sum_h s_h^{m,q} w_h}{\sum_h \hat{I}_h^{m,q} w_h}\right]^2 \tag{13}$$

where $s_h^{m,C}$ is the standard deviation within stratum h (Cochran 1977, sect. 5.4-5.9).[18]

[18] As in the case of SRS, the sampling fraction correction has been skipped.

Following the optimal allocation per strata (i.e. proportional to the standard deviation), sample size in each stratum can be expressed as follows:

$$n_{STRS,h}^{m,q}(\alpha,\beta) = n_{STRS}^{m,q}(\alpha,\beta)\frac{s_h^{m,q} w_h}{\sum_h s_h^{m,q} w_h} \tag{14}$$

Expression (13) suggests that stratified designs can reduce the source of discrepancies among aggregation formulas, depending on the ability of the former to reduce the variance within strata by means of a clustering approach able to isolate the criteria used to define pricing policies.

3.3. Alternative temporal targets

In the simplified framework under hypotheses A and B, expressions (10)-(12) and (13)-(14) have been for the moment referred to a generic reporting month *m*. They fix the number of product-offers whose price must be collected in the price reference month (*December y-1*) and in the generic reporting month *m* in order to achieve the desired precision level for the estimate of the price index in *m*. Nevertheless, if our aim is to produce complete annual series of monthly estimates, then a number of consequences do emerge, depending on the way we approach this task. The sample size needed for the estimates referred to month *m* is in fact in general different from the one needed to arrange the same precision for another month *m'*. This happens because the nature of price dynamics possibly changes from month to month in a way which may depend on the specific demand and supply characteristics of each consumer market.[19] Monthly samples can differ substantially, especially in the case of seasonal goods or services.

In any case the price collection in *December y-1* provides the base for the annual link, and therefore its role is crucial for all the monthly estimates that we are targeting: if we target a minimum precision level in every month, the sample in this price reference month has to be drawn in order to satisfy the size requirements of all the twelve following months. In particular, in SRS designs, sample size in the price reference month must be equal to the maximum size needed in the twelve months:

$$n_{SRS}(monthly) = \max_m (n_{SRS}^m(\alpha,\beta)) \tag{15}$$

If on one side the price collection in *December y-1* is the largest one, it might be not necessary to activate a monthly price collection extended to all this sample for the entire sequence of twelve monthly estimates. It is in fact possible to modulate price collection according to the actual monthly needs based on expressions (10)-(12). If we know that in a given month the expected variability is very low and that the desired precision can be achieved with a sample which is half the one drawn for the base according to (15), than we

[19] We may have to do with a very heterogeneous set of consumer markets - seasonal products, highly competitive markets, oligopolistic or monopolistic markets, markets highly dependent on external influences (markets for international commodities, weather, natural events) or even administered prices, and so on -, all behaving in quite different ways and with a variety of pricing policies. For a classification of price index dynamics within the HICP see De Gregorio (2011).

can save resources for price collection and concentrate them, for example, for the most critical months. A modular approach to price collection is then possible, and it gives the possibility to change the sample size every month in order to assure a given precision target in presence of heterogeneous variability patterns observed across months. In the case of seasonal products, for instance, this approach requires the largest effort in price collection in the price reference month, a high activation rate of the sample during peak months and lower off-peak rates.[20]

With stratified designs some further complications may arise since allocation is also a relevant factor. The sample size in the price reference month derives, in fact, from the sum of the largest monthly size of each stratum:

$$n_{STRS}(monthly) = \sum_{h} \max_m (n_{STRS,h}^{m}(\alpha,\beta)) \tag{16}$$

which may be much larger than the maximum overall monthly size derived from expression (14). This happens, in particular, whenever peaks in variability have distinct time patterns across strata, such as in markets characterized by seasonal pricing where peak months generally show higher variability: the timing of seasonal peaks, in fact, might differ across strata and this mere fact induces the need of larger samples in the price reference month. Something similar might happen in sectors characterized by highly irregular patterns.

The sub-indices with a relatively large variability or those characterized by seasonal behaviours are indeed only a part of the whole set of HICP sub-indices. It has been estimated that within the euro zone between 2004 and 2008 about 25% of HICP four-digit sub-indices showed a relatively strong monthly dynamics whilst about 7.3% showed a clear seasonal pattern (De Gregorio 2011). For what concerns the remaining indices, they were referred to markets where, price changes were quite regular and very slow, at least in periods of low inflation. In such cases, in the first months of the year – which are nearer to the base of *December y-1* - most observations are concentrated in the "no-change" zone: as a consequence, the distribution of price changes in those months is positively skewed. This asymmetry progressively loses ground as one moves away from the price reference month towards the final part of the year. If on one side the inertia of price indices in the first months reveals a very low variability and hence lower sample size needs, on the other side it might generate complications since the hypothesis of normality could not apply.

In any case, due to inertia, the last months of the year might be those in need of the largest samples, and the adoption of the annual link of December as a primary target for the estimates appears extremely reasonable: its importance relies in fact on the permanent effect that the link has on the chained index *H*.[21] In the case of the two types of design discussed above we obtain:

$$n(link) = n^{12}(\alpha,\beta) \tag{17}$$

This formula bears relevant gains in sample size with respect to expressions (15) and (16) only if the variability of price changes is diluted during the year and it is not

[20] For an application of this modular approach to seasonal products see De Gregorio, Munzi et al. (2008).

[21] See expression (1); Fenwick (1999) examines the issue of the choice of the price reference month, emphasizing the problems that may arise in the choice of the aggregation formula in case of large variability of price dynamics.

concentrated in the final month. Alternative reasonable targets might be set on quarterly or yearly averages:

$$n(quarterly) = \max_Q (n^Q(\alpha,\beta)) \tag{19}$$

and

$$n(yearly) = n^Y(\alpha,\beta) \tag{20}$$

In particular, for quarterly targets in stratified designs, allocation effects must also be considered as in the case of monthly estimates. It is also possible to use combined targets, for instance to guarantee the precision level on quarterly and annual link estimates.

4. Two case studies

4.1. Artificial populations

In order to test the combined effect of sample design and aggregation formulas on the variance of the estimates, we generated two artificial target populations starting from a selection of the microdata collected by ISTAT for the 2007 cycle of the HICP, and we iterated the extraction of samples from these populations in order to estimate the target parameter defined in Section 2. In particular, two case studies are here presented.[22] They are referred to price series characterized by high variability and heterogeneous behaviours: the first case regards European air transports, where the high volatility of price changes is partly explained by seasonal patterns; the other one regards package holidays, strongly affected by overlapping seasonal peaks with some inertia in the first months of the year.[23]

More formally, following the simplified approach outlined in par. 2.3, each set of microdata is interpreted as if it was a random sample drawn from the product-offers available in year y ($Z \subset \Omega^y$). Each record is characterized by a product identifier (g_i) and by a vector of 13 price quotes - from month *0* (the price reference month, namely December 2006) to month *12* (December 2007). For each market, a detailed and exhaustive partition of the goods in M_0 disjoint sets of consumption segments is then given:

$$\Gamma_{M_0} = \{G_h, h = 1,\ldots,M_0\}.$$

[22] Official microdata have been treated here with a different purpose from that pursued by ISTAT; it follows that results cannot be compared at any rate with the official figures currently disseminated.

[23] Flights and package holidays are both identified in De Gregorio (2011) as the sub-indices with the most heterogeneous behaviours across the countries of the euro zone, possibly needing further harmonization. For a methodological overview of the methods actually adopted by ISTAT to estimate these indices, see ISTAT (2009) and De Gregorio, Fatello et al. (2008).

Alternative but less detailed partitions Γ_{M_i} might be obtained by hierarchical aggregation of the subsets of Γ_{M_0}. For each partition a vector of normalized weights is accordingly defined:

$$W_{M_i} = \left\{ w_h, h = 1,..., M_i \mid \sum_h w_h = 1 \right\}.$$

Microdata in each set Z actually derive from stratified samples which have not been selected with probabilistic rules (ISTAT 2009; De Gregorio, Fatello et al. 2008, p. 20, 28-32). Nevertheless, they are treated here as if they were derived from random selections and the element of each set are expanded proportionally to the weighting structure W_{M_0} in order to form an infinite population. K simple random samples, each of n product-offers, are finally drawn from these infinite artificial populations. The yearly series of the monthly estimates $\hat{I}_k^{m,q}$ ($k=1,...,K$) are derived from each sample, adopting alternatively the Jevons, Dutot or Carli aggregation (expressions (7)-(9)). An inductive estimation of the sample mean variance is then produced and, consequently, an estimate of the sample size by means of the formulas derived in the preceding sections is provided. An identical approach is used to estimate the sample size for stratified designs based on alternative partitions of the target population.

All the simulations for the markets under scrutiny have been made by extracting iteratively 300 samples of 500 products each. Given a 1% error and a 95% confidence level, distinct temporal targets have been separately considered. Tables 1 and 2 (see par. 4.4 below) describe a relative measure of the sample size calculated as a multiple of a benchmark size (the one needed to estimate the yearly average with Carli aggregation and SRS). In particular the sample sizes have been determined in order to obtain the desired precision level for alternative temporal targets: i.e. separately for each single month, the cumulative target extended to the whole set of months (adopting expression (15) and (16)), the quarterly and yearly averages (expressions (18) and (19)). The desired precision target has been finally set on the link month of December, which - given the chaining procedure - affects permanently the fixed base series (expression (17)).

For the construction of the artificial population, in the case of European air transports we have used data from the original sample of $N=328$ product-offers, concerning as many European return flights connecting the country of origin (national) with the other countries (foreign). Each return flight is defined by a national and a foreign airport area (for instance, Rome and Frankfurt).

Four distinct partitions are used to provide alternative exhaustive segmentations of the target population. An elementary stratification Γ_{51} (51 strata) establishes an exhaustive segmentation by national and foreign regions (sub-national) and by type of carrier (low cost vs. full service carriers). A less detailed partition collapses the regions of each foreign country (Γ_{38}); a further aggregation of consumption segments uses only the country of destination and the type of carrier (Γ_{15}), and the less detailed partition only the country of destination (Γ_{11}). An elementary consumption segment can identify, for example, the low cost flights from the region A1 in country A (national) to the region B1 in the foreign country B; less detailed partitions identify, orderly, all the low cost flights from region A1 to B, all the low cost flights from A to B and all the flights from A to B.

In the case of package holidays[24] we have used the data from an original monthly sample of $N=246$ records. Each package is defined by a region of destination. Two distinct partitions provide exhaustive segmentations of the target population. A more detailed stratification Γ_{43} splits the universe into countries and type of holiday (i.e.: sea, mountain, city, etc.), while a less detailed segmentation Γ_{12} adopts only the splitting by country. As an example, an elementary segment could be the market for package holidays for type of holiday A1 in country A; a less detailed partition would concern all the packages for holidays for country A.

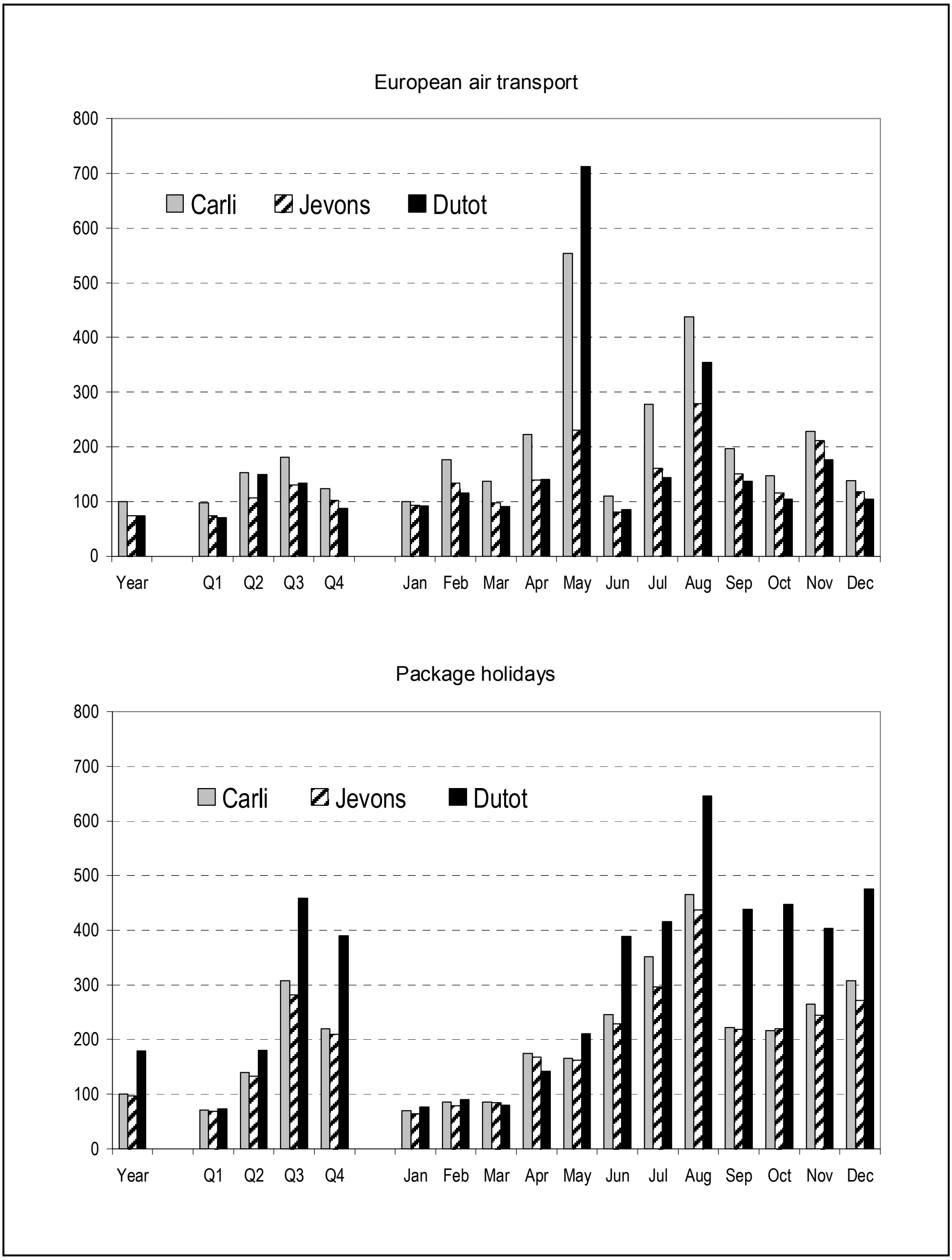
4.2. Design effect for independent temporal targets

Chart 1 plots the sample sizes needed with SRS for each month and quarter, and for the yearly average (in all the charts and tables, the sample size needed to estimate the yearly average with SRS and Carli aggregation is used as a benchmark and has been set equal to 100). Both markets show large differences in the variability within each month; Jevons formula delivers the best performance and Carli the worst, although heterogeneity in price levels seriously impairs the performance of Dutot aggregation during seasonal peaks; quarterly and yearly targets are far less demanding, although inertia effects may require larger samples in the last quarters.

In particular, for the separate estimates of the monthly indices of air transports smaller samples are needed at the beginning and at the end of the year (and in June) whilst the largest sizes are found in association with seasonal peaks in May and August (several times higher then the benchmark): in these months, in fact, the distributions of both price levels and price changes are positively skewed. Jevons aggregation is relatively less demanding, since it requires in May a sample size laying between two and three times the benchmark; in the same month the Dutot formula delivers by far the worst result (seven times the benchmark). Carli aggregation needs the largest size in eleven months out of twelve (with the median monthly size more than 35% higher than Jevons'). Dutot generates lower sample sizes in most of the off-peaks months (first and fourth quarter) due to a more appreciable homogeneity in price levels.

[24] Only foreign travels were considered.

Chart 1 - Sample size with SRS by sub-index, temporal target and type of aggregation (Indices. Base: size for yearly target with SRS and Carli aggregation = 100)



Moreover, although in air transports strong seasonal fluctuations hide any effect related to the time-distance between the reporting and the reference months, in the case of package holidays seasonal and inertia effects are combined: after the summer peaks (in July and August) sample size remains in fact quite large as compared to the first months of the year. This is the effect of the inertia of price dynamics, since price levels in the first months tend to range closer to their reference level and the estimates are less challenging. The annual link, in particular, seems to need a large sample as opposed to air transports where the link month was one of the easiest targets. In package holidays the worst performance is provided by Dutot aggregation: it works relatively well at the beginning of the year, but soon becomes by far the less appropriate (in terms of sample size) in the remaining months. This is due, probably, to the high heterogeneity of price levels, since they vary considerably across markets and show some likely correlation with price changes.

The figures for quarterly targets partially confirm this picture, although they are quite smoother for air transports where sample size never doubles the benchmark: Q1 requires the same sample as January or the yearly average separately for each aggregation method; Q2 and Q3 are more demanding, although they never double the sample size needed to estimate the yearly average. Package holidays on the contrary demand larger efforts in the last two quarters, due to the inertia effects, and confirm the inadequacy of Dutot aggregation, while Carli and Jevons require nearly the same sample size for all the quarters and for the yearly average.

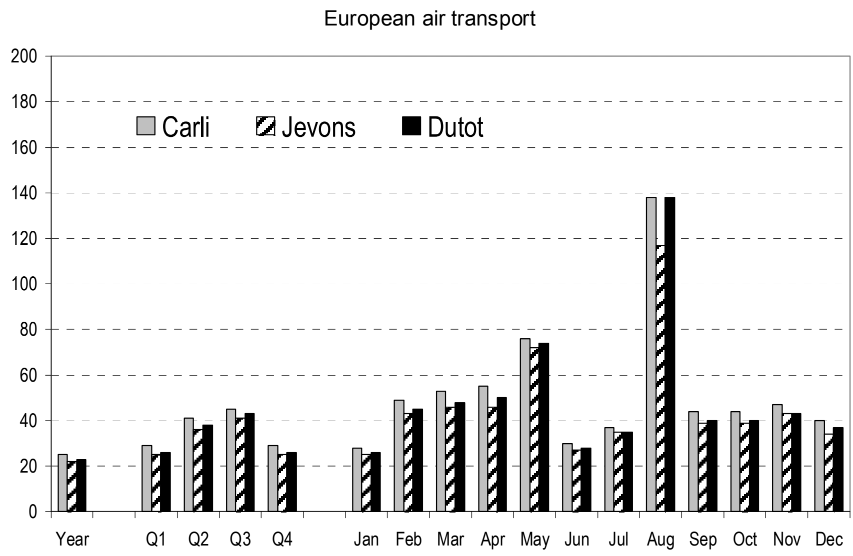
Chart 2 reports the effects on sample size deriving from the adoption of STRS at the most detailed level of stratification.[25] The effects of stratification are quite impressive: sample size is strongly reduced, the seasonal peaks are considerably smoothed and the differences among aggregation formulae tend to disappear.

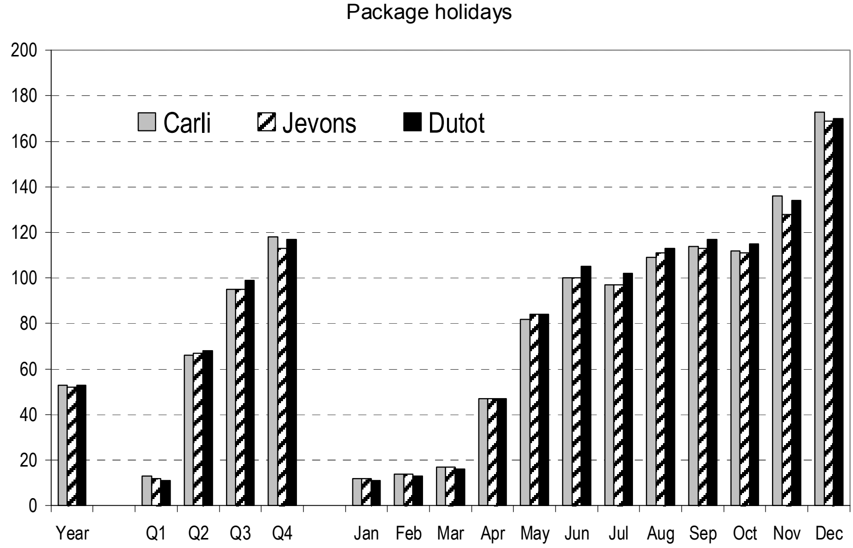
In air transports the sample size necessary to meet the yearly target is slightly more than 20% higher than the benchmark; quarterly samples and monthly samples are strongly reduced to 25-30% of the corresponding need in a SRS frame. Such decrease is particularly strong in peak months, especially in May. The effect of stratification is stronger with the Carli formula where the performance in terms of sample size improves considerably (sample size is only 10% higher than Jevons, in median). With the introduction of stratification, Jevons aggregation is still the one systematically requiring smaller samples: nevertheless, with stratification the differences in sample size due to alternative approaches to aggregation tend to shrink considerably.

This particular aspect is also evident in package holidays; stratification removes seasonal effects and only inertia plays a major. For the first three months samples are less the 20% of the benchmark, in June they pass 100% and the link month is the more demanding (nearly 170%). Quarters behave similarly, and this effect plays a key role in sustaining also the size of the sample required to meet the yearly target.

[25] Please, notice the different scale of this chart as compared to Chart 1.

Chart 2 - Sample size with STRS by sub-index, temporal target and type of aggregation (Indices. Base: size for yearly target with SRS and Carli aggregation = 100)





4.3. Combined temporal targets and modular price collection

In the previous sections, we discussed monthly or quarterly targets where each time span was considered independently from any other. What is more interesting is to see how large the sample has to be in order to meet at the same time all the monthly targets or all the quarterly targets. For SRS the solution is trivial: it is in fact sufficient to adopt the maximum monthly or quarterly sample sizes. On the contrary, with stratified designs strata allocation effects might complicate the matter (see par. 3.3): the relative efficiency of Jevons aggregation looses part of its advantage as compared to Dutot and Carli when combined targets are pursued, since Jevons allocation tends to show a higher heterogeneity in the sample size needed each month in each stratum. It can be said that the adoption of combined targets and stratified designs brings towards a reduction in the differences in efficiency due to the aggregation method.

More specifically, for the whole set of monthly targets of air transports with SRS the use of a Carli aggregation would need nearly 5.54 times the benchmark (Table 1). The Dutot formula delivers an even worse result (7.12), due to the high heterogeneity in price levels. The Jevons approach (2.79) needs half the sample size as compared with Carli. Such large samples derive from the high volatility observed for price levels and indices in peak months. If we reduce the SRS target to quarterly estimates, the sample sizes shrink drastically (between 1.30 to 1.81 times the benchmark) and the differences among the methods also are strongly reduced. The yearly estimates need nearly half the sample used for the quarterly target, while the annual link of December is placed between the quarterly and the yearly target. As long as the yearly target is concerned, Dutot equals Jevons' performance.

The effects of stratification are confirmed for combined targets: sharp reduction in sample size and more homogeneous results across the three aggregation approaches. The introduction of the first two levels of stratification brings large improvements, in particular for the Carli formula. The partition in 38 strata is extremely fruitful for all types of formulae, while the most detailed partition brings a comparatively minor reduction in sample size. In the passage from SRS to the most detailed stratified design there takes place a reduction of almost 70% of the necessary sample size. The temporal patterns of variability within strata are quite differentiated across months: consequently, the allocation effect induces appreciable differences between the sample size needed to target the whole set of monthly prices and the maximum size for separate monthly targets. If we consider the whole set of monthly targets (see Chart 1 and formula (16)), the sample size for Carli and Dutot aggregation is nearly 20% higher than the maximum size shown in Chart 1; Jevons formula, although it is in general more efficient, needs a sample nearly 30% higher than the respective maximum (151 vs. 117). For the quarterly indices, the size increase needed to meet all the monthly targets is slightly above 20%. The yearly target, independently of the design, requires about 15-20% of the sample size needed for the monthly targets, and the link demands nearly 30%.

In general Jevons aggregation performs better, with some exceptions where Dutot appears less demanding. Carli generally implies larger samples, although the differences collapse as stratification runs deeper. The heterogeneity of price levels damages the performance of the Dutot formula, especially where stratification is absent or limited, while the sample size derived from the Jevons formula appears less influenced by the presence of larger prices. If we consider package holidays, the irregularity of the monthly variability within strata appears less pronounced as compared to air transports (Chart 1). The SRS sample size needed to target the whole set of twelve months is in fact only 5% higher than the maximum size needed to meet separately the monthly targets (Table 1): Carli and Jevons aggregations require similar sample sizes (respectively, 4.65 and 4.37 times the benchmark), while Dutot has the worst performance. Most of the gains are obtained with the first level of stratification (12 strata): the three methods are almost equivalent. The adoption of further stratification confirms this picture and strongly reduces size needs. The equivalence of aggregation methods induced by stratification appears even stronger then in the case of air transports. It is worth the while to notice that, in any case (and differently from air transports), the estimate of the annual link for package holidays is more demanding than the estimate of quarterly and yearly targets, and that as stratification is adopted the estimate of the annual link requires a sample size which very near to the one needed to estimate the whole set of monthly targets.

Table 1 - Sample size, by sub-index, temporal target, type of aggregation and sample design (Indices. Base: size for yearly target with SRS and Carli aggregation = 100)

DESIGN	Aggregation	Temporal target			
		Monthly	Quarterly	Yearly	Annual link
		EUROPEAN AIR TRANSPORT			
	Carli	554	181	100	138
SRS	Jevons	279	130	74	118
	Dutot	712	149	74	104
	Carli	329	149	81	105
STRS 11	Jevons	239	105	59	81
	Dutot	284	116	63	88
	Carli	280	106	61	90
STRS 15	Jevons	225	90	50	74
	Dutot	267	98	55	81
	Carli	191	65	32	55
STRS 38	Jevons	171	59	28	48
	Dutot	188	61	29	50
	Carli	169	55	25	40
STRS 51	Jevons	151	49	22	34
	Dutot	167	52	23	37
		PACKAGE HOLIDAYS			
	Carli	465	307	100	307
SRS	Jevons	437	282	97	271
	Dutot	646	458	179	476
	Carli	258	183	78	245
STRS 11	Jevons	257	182	77	234
	Dutot	254	179	77	241
	Carli	183	123	53	173
STRS 43	Jevons	184	123	52	169
	Dutot	182	123	53	170

The pursuit of monthly targets might be very expensive and very much influenced by a few peak months. For this reason it might be redundant to extend price collection to the whole sample every month. It is instead possible to adopt cost-effective solutions based on modular approaches to price collection: only part of the sample might be surveyed every month, depending on the size needs imposed by the peculiar characteristics of variability in that month. The adoption of a modular month-dependent size to meet monthly or quarterly targets implies in any case a larger sample in *December y-1* – whose size can be derived from Table 1 - and reduced price collection especially in the months or quarters where the variability is lower.[26]

Table 2 reports some evidence. As we have seen, in the case of European air transports with SRS and monthly targets, the size index of the sample needed in the price reference month would be 554 in the case of Carli aggregation. Anyway, in the overall period of 13 months to December *y,* price collection can be skipped for more than a half (54.4%) of that sample. The saving is even higher in the case of Dutot aggregation (68%), while Jevons aggregation brings to a relatively lower saving (42.3%). Something similar happens for air transport when SRS is applied to quarterly targets, although in this case the sample sizes in the price reference month are much more homogeneous: average saving is around 20%, slightly for Dutot and Jevons. Stratification, as we have seen before, reduces the differences among aggregation methods and also the savings in price collection are quite similar: two units of the base sample out of three are saved on average with monthly targets and one out of three with quarterly targets.

Table 2 - Sample size reduction for price collection with modular sampling, by design, temporal target and type of aggregation (Indices. Base: size for yearly target with SRS and Carli aggregation = 100)

TEMPORAL TARGET PRICE COLLECTION	SRS			STRS (a)		
	Carli	Jevons	Dutot	Carli	Jevons	Dutot
			EUROPEAN AIR TRANSPORT			
Monthly target						
December y-1	554	279	712	169	151	167
Average 13 ..	252	161	228	62	55	59
Saving (%)	54,4	42,3	68,0	63,2	63,4	64,5
Quarterly target						
December y-1	181	130	149	55	49	52
Average 13 ..	142	105	113	37	33	35
Saving (%)	21,7	19,0	23,8	31,8	33,0	33,3
			PACKAGE HOLIDAYS			
Monthly target						
December y-1	465	437	646	183	184	182
Average 13 ..	300	279	453	123	122	125
Saving (%)	35,4	36,1	30,0	32,8	33,8	31,5
Quarterly target						
December y-1	307	282	458	123	123	123
Average 13 ..	194	181	289	77	76	78
Saving (%)	36,9	35,7	36,8	37,8	38,4	37,0

(a) Only the most detailed stratifications are considered here, i.e. 51 strata for European air transport and 43 strata for package holidays.

[26] Modular sample sizes are forcedly adopted in some specific markets, like in the case of accommodations in sites characterised by a strong seasonality (De Gregorio, Munzi et al. 2008).

Differently, in the case of package holidays with a SRS design Dutot aggregation is not only the less efficient method but also the one with the lowest saving deriving from the modular approach. With both SRS and STRS designs, slightly larger savings are obtained for the estimates of quarterly targets (between 35% and 40%).

Concluding remarks

This work has investigated and empirically tested some aspects of sample designs derived from the application of the most recent advancements in HICP methodology, by assuming a static definition of the target consumer market where replacements and changes in the range of products are excluded (see par. 2). As a whole, HICP concepts and methodology appear very well suitable for a more explicit use of the concepts and tools of statistical inference to estimate consumer price indices and to evaluate the quality of the estimates: for this reason, they also pave the way for a more cost-effective planning the technical management of monthly surveys.

In particular, we have analysed and compared the sample sizes requirements by combining the adoption of simple or stratified random sampling with alternative approaches to elementary aggregation and with a set of temporal targets. As a first step, we derived in par. 3 the expressions for a generic monthly sample size by type of design and aggregation, and developed them as functions of the coefficient of variation of indices and price levels: our findings confirm that aggregation effects on optimal sample size depend crucially on the level of relative variability and skewness of observations; such effects tend to annul when price changes are smoother and if the precision target is sufficiently tight.

The case studies reported in section 4 confirm these results and highlight some more points: the crucial role of stratification in saving sample size; the heterogeneity of the results obtained with different approaches to aggregation, and its fading out as stratification is introduced and when allocation effects are at work in stratified designs with multiple temporal targets; the possibility to adopt modular schemes of price collection especially with strongly seasonal items; the options opened by fixing temporal targets alternative to the monthly series, especially quarterly averages or the annual link; the role of indices' inertia in the determination of the sample size of the annual link.

Empirical evidence shows that stratification may shrink sample size by 50% to 70% as compared to SRS design. The choice of the strata is of paramount importance, since it involves theoretical and microeconomic issues: here it has been based on marketing criteria, trying to isolate possibly homogeneous consumption segments and clusters of pricing policies. The issue of how deep stratification should be is also very important. The introduction of a first layer with a few strata brings immediately large gains in sample size. More complex stratifications usually - but not necessarily - produce comparable gains with respect to more elementary designs. This depends obviously on the relative efficiency of a deeper stratification to compress the variance within strata. In the case of air transports, for example, adding the type of carrier to the country of destination increases by nearly 40% the number of strata but does not seem to generate very large gains, at least with the Jevons or Dutot aggregation. On the contrary, more detailed areas of destination produce important size gains, since they probably better reflect the pricing criteria of this market.

The case studies all emphasise the role that market segmentation and stratification have in reducing optimal size, increasing precision and saving resources. Stratified designs can produce other interesting effects, such as reducing drastically the heterogeneity resulting from alternative aggregation methods. Very heterogeneous optimal sizes might in fact derive from Carli, Dutot or Jevons approaches, depending on the variability of price changes and price levels. It is well known that with no - or just with a few - strata, Jevons performs significantly better in terms of optimal size. Deeper stratifications tend anyway to reduce this advantage due to within-strata homogeneity. The choice of the aggregation method is an issue largely debated in literature but it loses importance as stratification is considered, especially with large and highly stratified samples. Even the adoption of combined targets produces some smoothing for the aggregation effect: empirical evidence suggests that Jevons aggregation loses part of its advantages when the target is moved from a single month to the whole series of monthly indices, due to a less favourable monthly allocation of the units across strata.

The fact that optimal sample size might be determined on a monthly basis has also a number of consequences. Even if we stick to a defined approach to sampling and aggregation, heterogeneous monthly results might occur in markets where the variability of pricing behaviours is monthly dependent. This is likely to happen with seasonal items or even in markets where the variability of prices is somewhat structural: the cases of flights and package holidays are paradigmatic. In such a context, adopting a constant monthly size in price collection appears sub-optimal. What we intend to highlight is that a modular approach to price collection is possible, allowing a concentration of resources in those months where variability hits a peak, and consequently favouring a better management and scheduling of the surveys. Empirical evidence suggests the adoption of a modular price collection, with strong efforts concentrated in the price reference month while part of the monthly samples can be drastically reduced.

In this respect, the consideration of alternative temporal targets also appears as a strategic issue, if one of the objectives is to save resources by optimising their use. When quarterly targets are concerned, considerable gains in sample size are obtained as compared to monthly targets: large differences among aggregation methods anyway persist also if the target is moved on the yearly average or on the annual link and unless highly stratified designs are considered. Targeting the annual link is justified by its permanent effect on the chained index: such target may imply a large gain in sample size, as it happens for air transports; but if the link month is among those showing a higher variability (as in the case of package holidays) this objective may not produce large enough gains.

Although this work is based on several restrictive hypotheses on the dynamics of the set of the available product-offers (time invariance, with no changes in the range of the products and in the retail network), such hypotheses were essential in order to provide a reliable definition of the statistical target and a one-to-one mapping for the re-pricing of the set of the transactions in the weight reference period based on the product-offers available in the price reference and in the reporting month. This can be interpreted as a first approximation: relaxing these hypotheses implies in fact a huge modelling of consumers' choices in order to produce more sophisticated mapping functions. Further developments on these issues might be obtained both on the theoretical and empirical grounds. Concerning the first, the pioneering work of Ribe (2000) deserves more analysis on the form and nature of the mapping functions and their implications, especially with reference to the structural characteristics of consumer markets. It could be fruitful to consider different classes of mapping functions to be used in particular clusters of consumer markets. Empirical studies might help in this work, by examining other sectors and by providing deeper insights on the relative efficiency of alternative stratification criteria. Under this respect, the ambiguity of consumption segments and the role of stratification need also further empirical research, especially for what concerns the study of supply and demand effects on specific consumer markets on a case-by-case basis in order to isolate the sources of pricing behaviour.

A further remark concerns the weighting strategy. In this paper it was assumed that weights are not a source of potential statistical error, although weighting are estimates themselves and are a primary source of error, being often at the core of the criticism against official CPI estimates. Nevertheless, a specific and structured literature on the subject is lacking (remarkable exceptions, such as Biggeri et al. 1987, do not impair this statement), although the adoption of confidence intervals and precision targets cannot ignore this issue. Work on this subject is thus necessary, with the objective to join together the effects of price and price indices variability with those of weights variability. Finally, the role of overall inflation has also to be considered: expected variability of price levels and price changes is strictly connected with the expected evolution of inflation expectations. This aspect also should be modelled in order to achieve a more complete approach to CPI sampling: quite surprisingly, also in this case literature is lacking.

References

Adelman, I. 1958. A new approach to the construction of index numbers. *Review of economics and statistics*, 40, p. 240-249.

Andersson C., G. Forsman, J. Wretman. 1987. On the measurement of errors in the Swedish consumer price index. Invited paper 12.2 at the 46^{th} session of the ISI-IASS meeting, Tokio 8-16 September. Booklet, p.261-277.

Balk, B.M. 2003. Price indexes for elementary aggregates: the sampling approach. Paper presented at the seventh meeting of the International working group on price indices, Paris, 27-29 May.

Balk, B.M. 2008. *Price and quantity index numbers. Models for measuring aggregate change and difference*. New York: Cambridge university press, p.283.

Banerjee, K.S. 1956. A note on the optimal allocation of consumption items in the construction of a cost of living index. *Econometrica*, Vol. 24, No. 3 (July), p. 294-295.

Baskin, R.M., S.G. Leaver. 1996. Estimating the sampling variance for alternative forms of the U.S. consumer price index. *Proceedings of the Survey research methods section, American statistical association*, p. 192-197.

Biggeri, L., P.D. Falorsi. 2006. A probability sample strategy for improving the quality of the consumer price index survey using the information of the business register. Paper presented at the joint ECE/ILO meeting of the of Group of experts on consumer price indices, Working paper n.12, Geneva 10-12 May.

Biggeri, L., A. Giommi. 1987. On the accuracy and precision of the consumer price indices. Methods and applications to evaluate the influence of the sampling of households. Invited paper 12.1 at the 46^{th} session of the joint ISI-IASS meeting, Tokio 8-16 September. Booklet, p.244-260.

Cochran, W.G. 1977. *Sampling techniques*. 3rd edition. New York: John Wiley and sons, p. 428.

Dalèn, J. 1992. Computing elementary aggregates in the Swedish consumer price index. *Journal of official statistics*, Vol. 8, No.2, p. 129-147.

Dalèn, J. 1998. On the statistical objective of a Laspeyres' price index. Paper presented at the fourth meeting of the International working group on price indices, Washington D.C., 22-24 April.

Dalèn, J. 2001. Statistical targets for price indexes in dynamic universes. Paper presented at the sixth meeting of the International working group on price indices, Canberra, Australia, 2-6 April.

Dalèn, J., E. Ohlsson. 1995. Variance estimation in the Swedish consumer price index. *Journal of business & economic statistics,* Vol.13, No. 3 (July).

De Gregorio, C. 2011. The variability of HICP sub-indices and harmonisation needs. *Rivista di statistica ufficiale*, n.1 p. 5-32.

De Gregorio, C., S. Fatello, R. Lo Conte, S. Mosca, F. Rossetti. 2008. Sampling design and treatment of products in ISTAT centralised CPI surveys. *Contributi ISTAT,* n. 1.

De Gregorio, C., C. Munzi, P. Zavagnini. 2008. Problemi di stima, effetti stagionali e politiche di prezzo in alcuni servizi di alloggio complementari: alcune evidenze dalle rilevazioni centralizzate dei prezzi al consumo. *Contributi ISTAT*, n. 6.

De Haan, J., E. Opperdoes, C. Schut. 1997. Item sampling in the consumer price index: a case study using scanner data. Paper presented at the joint ECE/ILO meeting of the Group of experts on consumer price indices, working paper n.1, Geneva 24-27 November.

Dippo, C.S., K.M Wolter. 1983. A comparison of variance estimators using the Taylor series approximation. Proceedings of the Survey research methods section, American statistical association, p. 113-119.

Dorfman, A.H., J. Lent, S.G. Leaver, E. Wegman. 2006. On sample survey designs for consumer price indexes. *Survey methodology*, Vol. 32, No. 2 (December), p. 197-216.

EUROSTAT. 2001. *Compendium of HICP reference documents*. Bruxelles: European communities.

EUROSTAT. 2003. Outline of a Possible strategy for harmonizing HICP lower level aggregation. Paper presented at the meeting of the working party "Harmonization of consumer price indices", Luxembourg, 9-10 December.

Fenwick, D. 1998. The impact of sample design on the performance of the sample geometric mean and related issues. Paper presented at the fourth meeting of the International working group on price indices, Washington D.C., 22-24 April.

Fenwick, D. 1999. The impact of choice of base month on the relative performance of different formulae used for aggregation of consumer price index data at an elementary aggregate. Paper presented at the fifth meeting of the International working group on price indices, Reykjavik, 25-27 August.

Fenwick, D. 2001. Sampling in consumer price indices: what role for scanner data? Paper presented at the sixth meeting of the International working group on price indices, Canberra, 2-6 April.

Hansen, C.B. 2006. Price updating of weights in the CPI. Paper presented at the ninth meeting of the International working group on price indices, London, 14-16 May.

Koskimäki, T., M. Ylä-Jarkko. 2003. Segmented markets and CPI elementary classifications. Paper presented at the seventh meeting of the International working group on price indices, Paris, 27-29 May.

Kott, P.S. 1984. A superpopulation theory approach to the design of price index estimators with small sampling biases. *Journal of business & economic statistics*, Vol.2, No.1 (Jan).

ILO (ed.). 2004. *Consumer price index manual: theory and practice*. Geneva

ISTAT. 2009. Come si rilevano i prezzi al consumo. www.istat.it.

Leaver, S.G., J.E. Johnstone, K.P. Archer. 1991. Estimating unconditional variances for the U.S. consumer price index for 1978-1986. Proceedings of the Survey research methods section, American statistical association, p. 614-619.

Leaver, S.G., W.L. Weber, M.P. Cohen, K.P. Archer. 1987. Item-outlet sample redesign for the 1987 U.S. consumer price index revison. Invited paper 12.3 at the 46th session of the ISI-IASS meeting, Tokio 8-16 September. Booklet, p.278-286.

Leifer, H.-A. 2002. The ‘elementary level’ of a consumer price index and cost-of-living index: similarities and differences. Paper presented at the meeting of the Working party "Harmonization of consumer price indices", Luxembourg, 30 September-1 October.

Leifer, H.-A. 2008. Are the Carli, Dutot and Jevons indices time-path dependent? Paper presented at the fifth meeting of the HICP Task force sampling, Luxembourg, 19 February.

Norberg, A. 2004. comparison of variance estimators for the consumer price index. Paper presented at the eight meeting of the International working group on price indices, Helsinki, 23-25 April.

Norris, N. 1940. The standard errors of the geometric and harmonic means and their application to index numbers. *The annals of mathematical statistics*, Vol. 11,No. 4 (Dec.), p. 445-448.

Ribe, M. 2000. Price index, statistical inference and harmonisation. Proceedings of the fourth International conference on methodological issues in official statistics. 12-13 October.

Silver, M., S. Heravi. 2006. Why elementary price index number formulas differ: price dispersion and product heterogeneity. *IMF working paper*, WP/06/174.

Viglino, L. 2003. Harmonisation of HICP lower level aggregation. The use of the geometric mean for EEG and EA aggregation. Paper presented at the meeting of the Working party "Harmonization of consumer price indices", EUROSTAT, Luxembourg, 9-10 December.

Wilkerson, M. 1967. Sampling error in the consumer price index. *Journal of the american statistical association*, Vol. 62, No. 319 (Sep.), p. 899-914.

Young Entrants, Temporary Jobs and Career Opportunities: Short-Term Perspectives of Young Italian Workers

Marco Lilla, Stefano Staffolani[1]

Abstract

Because of the high unemployment rate and the low chances to get a permanent job, young people in Italy face major problems in entering the labour market. In this paper, we address the issue of the labour market entry of young Italian workers by using a sample dataset of the new Compulsory Communication data for the period 2008 - 2010. We ask whether there are some individual characteristics which are more favourable for open-ended contracts. Then, we analyse how different types of temporary jobs have different probability of transforming into permanent ones. Finally, we estimate a survival model on the duration of temporary jobs.

1 Introduction

The last estimates from the labour force survey in Italy state that more than 1 out of 3 economically active individuals aged between 15 and 24 years are officially unemployed. Furthermore, the incidence of temporary contracts is higher for the younger cohorts and young workers exhibit very low initial wages. Berton et al. [2009a] show that the cost associated with some temporary contracts (wage and salary) can shrink to half the one of standard workers in Italy.

However, as worker security is interpreted as a way for firms to create and retain in-firm specific skills, employers do not necessarily have an interest in indiscriminate use of short temporary contracts, i. e. extensive flexibility.

The use of open-ended contracts induces more human capital accumulation, especially of the firm-specific form, because a higher expected duration of the employment spell (see Berton et al. 2009b) provides more incentives to both workers and firms to invest in training.

As for career expectations in Italy, the ambiguous nature of temporary contracts - on the one hand the opportunity of receiving training, on the other hand the risk of being trapped in temporary jobs - seems to be correlated with the duration of the contract.[2] The probability of transition to an open-ended contract decreases with the number of temporary

[1] Dipartimento di Economia, Università Politecnica delle Marche, m.lilla@univpm.it
Compulsory Communication data are used thanks to the agreement between our Department of Economic and Social Sciences and General Department for the Innovation Technology of the Ministry of Labour and Social Policies. We thank Grazia Strano, Giuliana Coccia and the participants to the 2011 Workshop for the Enhancement and Social Responsibility of Official Statistics held in Rome.

[2] On the contrary, in Britain temporary contracts were still a port of entry to non-temporary contracts during the ’90s (Booth et al., 2002.

experiences and the duration of unemployment spells, as shown in Gagliarducci (2005), who used the ILFI dataset (Indagine Longitudinale sulle Famiglie Italiane).

Moreover, Berton et al. (2009b), using the WHIP database (Work Histories Italian Panel), find that different types of temporary contracts lead to different impacts on the probability of obtaining an open-ended position. They also show a significant difference between temporary contracts with training ("contratto di formazione lavoro e apprendistato") with respect to freelance ones ("lavoro parasubordinato").[3]

According to Picchio (2008), who uses the SHIW database (Survey of Italian Households' Income and Wealth), in Italy workers employed on temporary basis have a higher probability of obtaining a permanent position than the unemployed ones, although there are state-dependence effects on being held in temporary positions.

It is worth noting that temporary contracts imply a higher volatility of wages over time. In Italy, the unbalanced generational wage dynamic stands out by comparing the average wage of different generations and their levels of instability (Staffolani et al., 2009).

The paper aims at evaluating the job market entry of the younger cohorts in Italy by using more than two years job histories from the new data from Compulsory Communications ("Comunicazioni Obbligatorie", CC hereafter). Since the beginning of 2008, each employer has to communicate some events related to his employees' contracts, namely activation, fixed-term extension, transformation and anticipated termination (see Strano et al., 2010).

At a first stage we ask whether there are any individual characteristics as well as employer features which are more favourable for obtaining permanent jobs.[4] Then, we follow cohorts of young individuals which entered the job market at the beginning of 2008, observing their career profiles in the following two years.

The next section presents the data. Section 3 analyses the characteristics of the entry and exit processes of young cohorts in the Italian labour market and presents a survival analysis on jobs duration. Section 4 concludes.

2 Data: from contracts to jobs

The sample dataset comes from the CC system and refers to the events for all the workers born on 15 of March, June, September and December: about 1 out of 91 of all italian workers have been included in the CC sample.[5] In order to avoid some problems related to the early development of the system and to the failure to review the latest communications, CC have been selected for the period from April 2008 up to June 2010.

Initially, the sample dataset consisted of more than 303.000 contracts. However, temporary contracts which started before 2008 and ended in the following period at the due date of termination are not included. Therefore, the labour contracts terminations are

[3] It is a peculiar Italian labour contract which presents both the characteristics of the employment contract and the self-employment.

[4] Permanent jobs are open-ended contracts, while temporary jobs refer to contracts with a scheduled ending date. In our data we observe five different types of temporary contracts, i.e. fixed-term, apprenticeship, freelance, interim and internship.

[5] Since January 2008 each employer, both in the public and the private sector, must communicate via the website *http://www.co.lavoro.gov.it/* his employees' contracts. The CC database does not include self-employed.

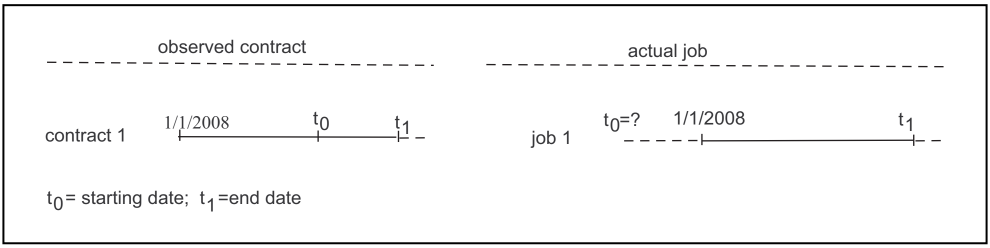
underestimated and the analysis of the differences between activations and terminations of labour contracts - inflows and outflows - is misleading at the moment.

The discussion about our CC sample dataset needs a preliminary caveat. While the CC dataset refers to contracts, identified by a starting date of a job appointment between an employer and an employee, the one we used in the following analyses has been modified to focus on jobs, characterized by a continuous relationship - or almost continuous, as we will see below - between the same employer and employee. In order to move from contracts to jobs, we carried out the following steps:

- there are many cases of transformation, extension, early termination of contracts with a starting date after 2008, nevertheless their activation is missing even if the registration should have been compulsory. In these cases, we can assume that employers probably incorrectly specified the starting date of the contract: the job should be started before 2008. For our purposes, the starting date of these contracts has been substituted with a missing value - see figure 1 - ruling out these jobs from the analyses (about 9.000 contracts);
- many temporary contracts have been extended once or more with no interruption. A unique job has then been considered with the final expiry date equal to the termination of the last extension, keeping the counts of the extensions;
- sometimes the employer takes the employee back after a short period from the end of the previous temporary contract. If the time between the ending date of the first contract and the starting date of the second one is lower than a given number of days, namely τ,, these cases represent "hidden extensions" (see Figure 2);
- finally, two different contracts (for instance, fixed-term and open-ended) between the same worker and firms actually could represent "hidden transformation", when the ending date of the first contract and the starting date of the second one (the distance, τ) is lower than a given number of days (see Figure 3).

We set $\tau = 30$[6] for the cases of hidden extension and hidden transformation (about 47.000 cases). It is worth stressing that our jobs differ from the administrative definition of contracts, because we are interested in the almost continuous job relationship between a worker and a firm. Hence, one job can refer to several consecutive contracts.

Figure 1 - "Implausible" contract created by a communication



[6] The choice of a monthly period is made because this is the minimum length analysed in the empirical researches based on survey data, see Gomes (2012).

Figure 2 - "Hidden" extension

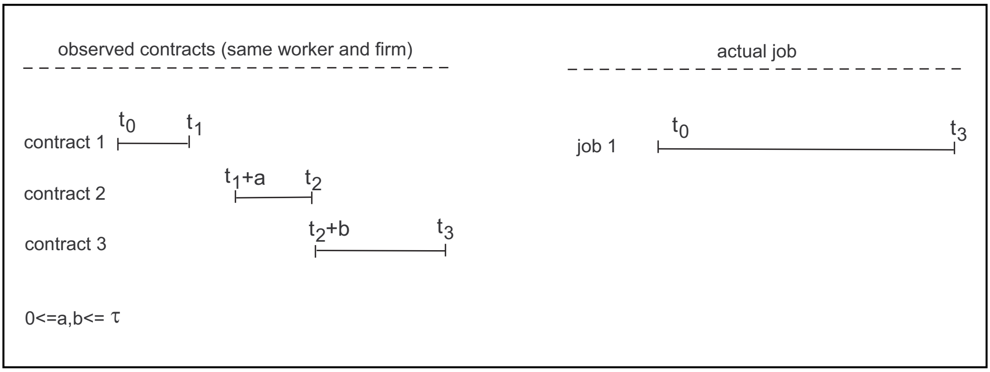
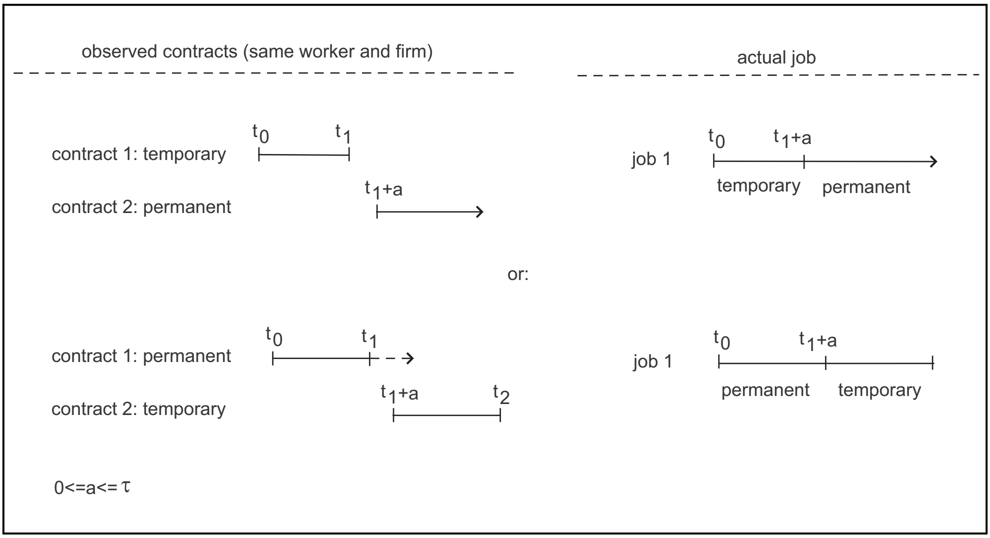


Figure 3 - "Hidden" transformation



Following the steps outlined above, we end up with the *jobs* employer-employee dataset. The analyses presented thereafter refer to *young workers*, defined as individuals born between 1974 and 1993, whose job started from April 2008 up to 30 June 2010, ending with 82.419 “jobs” included in the sample.

Note that our analyses refer to labour relationships and not to workers. When a worker has multiple jobs during the period, or she moves to another firm, or she leaves her job (or she is laid off) and later is newly hired by the previous employer (and unless the period between the two contracts is lower or equal to 30 days, see above), then there are multiple observations for the same worker.

3 Temporary and Permanent Jobs

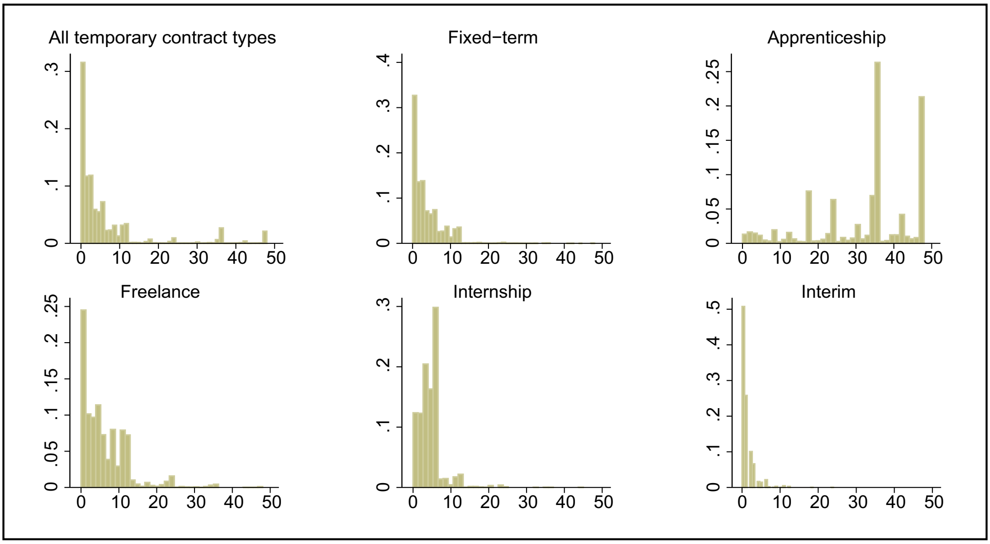
The average number of jobs for young workers is around 2.6 during the period that we considered, 60% of young workers shows only one presence in the dataset, 23.4% are present two times and only a small fraction had 3 or more jobs during the period. Table 1 shows the type of contract under which young people are hired. More than 3 out of 4 jobs start as temporary.[7]

Table 1 - Jobs by types of contract

	Jobs	%	Expected (ex-ante) average length, months
Permanent	22022	27	-
Fixed-term	39387	48	3,99
Apprenticeship	7030	9	37,20
Freelance	6507	8	9,33
Internship	2513	3	4,52
Interim	4960	6	1,32
Total	**82419**	**100**	**8,23**

The 25% of all jobs started in the second quarter of 2008 was permanent; the same figure went down to 19% in the same quarter of 2010. The expected lengths of jobs by type of contract[8] are shown in figure 4. Different types of contracts are characterized by very different distributions of the expected length. The median for the length of fixed-term and interim contracts is one month, whereas the other types of contracts have higher medians. The highest average length is for the apprenticeship contract (37 months) and the lowest is for the interim one (1.3 months). The average length of temporary contracts is equal to 8.2 months.

Figure 4 - Temporary contracts and their expected (ex-ante) length, months



[7] The same figure for older workers is less than 2 out of 3.

[8] In the graph were plotted those jobs with an expected length lower or equal to 50 months, which account for the 97% of temporary jobs. The "actual" length of temporary jobs takes into account that some temporary contracts have been transformed in permanent ones, while other contracts have been deferred or ended in advance.

The freelance contracts mostly regard workers with an academic degree, that are also the less engaged in apprenticeship. Permanent contracts are most associated with primary education as shown in table 2.[9]

Table 2 - Jobs distribution by types of contract and education, column %

	Primary	Secondary	Tertiary	Total
Permanent	25.7	20.3	17.3	22.5
Fixed-term	52.0	47.0	43.6	48.9
Apprenticeship	10.5	10.0	4.6	9.5
Freelance	4.1	11.4	20.6	9.1
Internship	2.3	3.7	8.7	3.7
Interim	5.6	7.6	5.3	6.3
Total	**100**	**100**	**100**	**100**

Permanent jobs concern men more than women and immigrants than Italians. Freelance jobs concern more frequently women and italians.

Table 3 - Jobs distribution by types of contract and gender/nationality, column %

	Male	Female	Immigrant	Italian	Total
Permanent	29.4	23.6	41.0	21.4	26.7
Fixed-term	46.7	49.0	45.5	48.6	47.8
Apprenticeship	8.9	8.0	6.2	9.4	8.5
Freelance	6.1	10.0	1.9	10.1	7.9
Internship	2.5	3.6	1.1	3.8	3.1
Interim	6.4	5.6	4.3	6.7	6.0
Total	**100**	**100**	**100**	**100**	**100**

Interesting differences emerge also by considering the geographical dimension: permanent jobs were used more frequently in the North-West, Islands and South while interim jobs are widely used in the northern regions as shown in table 4.

Table 4 - Jobs distribution by types of contract and geographical area, column %

	North-West	North-East	Center	South	Islands	Total
Permanent	29.9	21.2	25.7	28.5	29.5	26.7
Fixed-term	38.6	52.3	46.1	53.7	51.7	47.8
Apprenticeship	9.2	10.8	9.8	5.3	5.9	8.5
Freelance	9.0	5.2	9.5	7.2	9.4	7.9
Internship	3.5	3.6	3.6	2.1	1.6	3.1
Interim	9.7	7.0	5.4	3.1	1.8	6.0
Total	**100**	**100**	**100**	**100**	**100**	**100**

9 Some jobs refer to individuals with missing data on education, the 23% of cases, many of them being immigrants. Therefore, the row totals differ from the previous table.

In order to sum up the previous evidences and check their significance, a logit model has been estimated, where the dependant variable is the probability for the job to start as permanent versus temporary.

Table 5 - Logit estimates of getting a permanent job, youngs

	A1	A2	A3
	Odd ratios	Odd ratios	Odd ratios
Quarters, ref: 2008q2			
2008q3	0.87***	0.89***	1.06*
2008q4	1.08**	1.08**	1.07*
2009q1	1.29***	1.15***	1.16***
2009q2	1.31***	1.13***	0.86***
2009q3	0.71***	0.65***	0-79***
2009q4	0.95	0.89***	0.91**
2010q1	1.12***	0.96	0.98
2010q2	0.73***	0.65***	0.66***
Age		1.07***	1.09***
Female		0.83***	0.85***
Italians		0.50***	0.59***
Education, ref: compulsory			
Primary		1.44***	1.23***
Secondary		0.86***	0.87***
Tertiary		0.61***	0.80***
Area, ref: North-West			
North-east		0.60***	0.64***
Center		0.81***	0.77***
South		1.06**	1.41***
Islands		1.16***	1.26***
Other control variables:	NO	NO	Occupation; Sector
N	82416	82394	77377
Pseudo-R2	0.008	0.072	0.216
LR χ^2	725.1	6949.9	20001.5

Some differences emerge from the different specifications of the model. In the following comments, we refer to the full specification in the *A*3 column. The probability for the job to start as permanent increased during the first period of the crisis, whereas it significantly decreased from the mid 2009 and increases with age among young individuals,and is higher for males and foreigners. Lower educational attainments (the reference group) imply a higher probability for the job to start as permanent. With respect to the North-West area, firms located in the North-East and in Central Italy make a widespread use of temporary contracts.

Focusing now on the temporary jobs, they can end at their expiry date, or be extended or transformed into permanent ones[10] or end in advance. The above four different types of outcomes of temporary jobs are presented, for the different types of temporary contracts, in table.

[10] Remind that we consider also the hidden extensions and the hidden transformations, see section 2.

Table 6 - Outcomes of temporary jobs at their expected ending date, by types of contract

TEMPORARY JOBS	Ended at expiry date	Transformed into permanent	Extended	Ended in advance	Total
Fixed-term	46.1	7.9	25.2	20.8	100
Apprenticeship	43.5	5.3	2.3	49.0	100
Freelance	50.9	2.0	26.4	20.8	100
Internship	52.1	3.6	24.9	19.3	100
Interim	33.8	5.1	52.6	8.6	100
Total	**45.5**	**6.6**	**24.9**	**23.0**	**100**

The highest probability of transformation concerns fixed-term jobs, the lowest one concerns freelancers. Apprenticeship shows the highest probability of ending before the expiry date, whereas interim jobs show the highest one of extension.

According to our CC data, only 6.6% of temporary jobs are transformed into permanent ones. Given that such transformation is the most important change that can affect a temporary job, we present in table a multivariate descriptive statistic based on a logit model.[11]

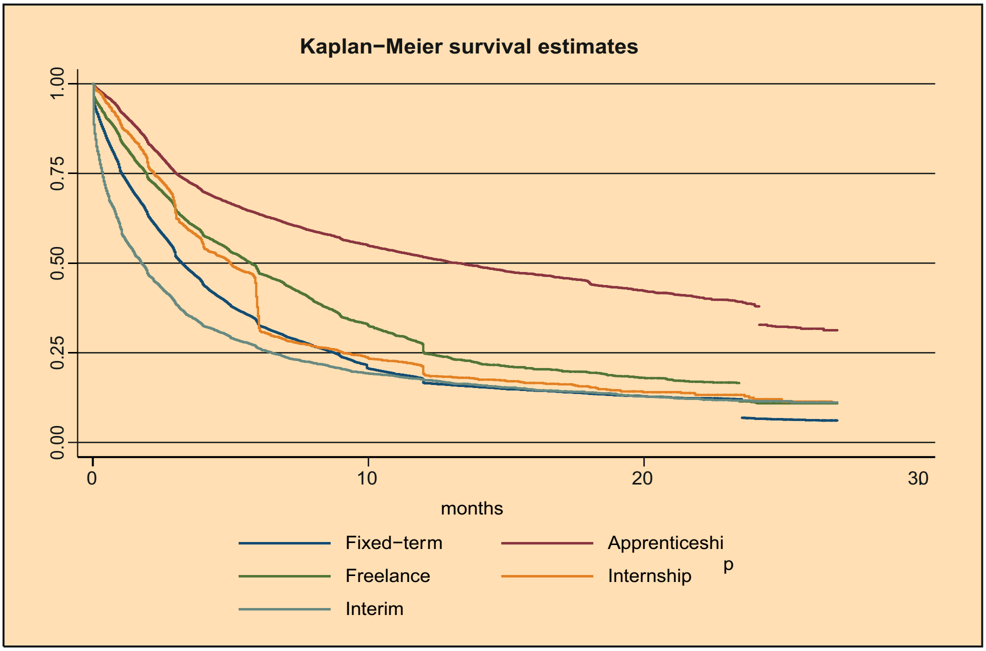
Table 7 - Logit estimates of transforming a temporary job into a permanent one, youngs

	B1	B2
	Odd ratios	Odd ratios
Quarters. ref: 2008q2		
2008q3	1.02	1.00
2008q4	0.94	0.93
2009q1	1.04	1.07
2009q2	0.78***	0.77***
2009q3	0.76***	0.75***
2009q4	0.71***	0.71***
2010q1	-0.81***	0.82**
2010q2	0.83*	0.82**
Age	1.07***	1.06***
Female	0.87***	0.86***
Italian	0.81***	0.82***
Education, ref: compulsory		
Not Available	0.98	0.98
Secondary	1.05	1.06
Tertiary	1.30***	1.49***
Plant area, ref: North-West		
North-East	0.77***	0.74***
Center	0.78***	0.79***
South	0.46***	0.47***
Islands	0.41***	0.42***
Temporary Contract, ref: Fixed-term		
Apprenticeship		0.67***
Freelance		0.16***
Internship		0.27***
Interim		0.13**
Other control variables:	Occupation; Sector	Occupation; Sector
N	41277	41277
pseudo-R2	0.131	0.156
LR χ^2	3275.6	3898.103

[11] We drop temporary contracts whose expected expiry date was later than 30/06/2010.

Starting from the second quarter of 2009, the probability of transformation of a temporary job into a permanent one strongly reduced. Jobs referring to women and Italian citizens show a lower probability of transformation as well as those concerning workers with educational attainments lower than tertiary education.. This probability is higher in the North-West, the reference area. The fixed-term contract is associated with a higher probability of transformation into a permanent one.

Figure 5 - Effective (ex-post) job duration (in months)



The main differences between expected (ex-ante) and actual (ex-post) durations respectively in figures 4 and 5 arise mainly from the different probabilities of being extended or ended in advance as shown in table. There are very different actual durations with respect to the types of temporary contracts (see Figure 5): apprenticeships are the most long-lasting while internships show the lower duration with major terminations after six months.

The CC dataset comes as a set of survival-time data where the subject is the job. Obviously, there is a censoring for the temporary jobs which should terminate after the end of the period we observe, June 2010.[12] Given this data structure, we can ask whether effective durations are associated to some individuals and job characteristics by means of a simple survival model. Table 8 shows the estimates for the hazard of the termination of a temporary job for some individual and job characteristics, controlling for time changes,

[12] These temporary jobs ending after 30th June 2010 are nevertheless included in the sample and treated as surviving ones as usual in survival models.

regions, occupations and sectors. Accounting for different types of temporary jobs may result in changes of the effects of the individual characteristics, i.e. there are significant differences across types of temporary jobs for the individual characteristics. Looking at the full specified model (column C2), women exhibit higher hazard rate for the termination of their temporary jobs - even if less significant - as well as for the Italian citizens. Age and education have a negative impact on the hazard, while the probability of termination increases from the North to the South Italy. With respect to the different types of temporary jobs, fixed-term jobs show the highest hazard of termination, and apprenticeships the lowest. These hazards have increased since the early 2009.[13]

Table 8 - Survival estimates of the hazard for a temporary job, workers 15-34

	C1	C2
	coeff.	coeff.
Quarters. ref: 2008q2		
2008q3	0.044**	0.010
2008q4	0.030	0.037*
2009q1	-0.193***	-0.189***
2009q2	0.106***	0.069***
2009q3	0.134***	0.056***
2009q4	0.074***	0.036*
2010q1	-0.212***	-0.203***
2009q2	-0.266***	-0.310***
Age	0.008***	-0.030***
Female	0.046***	0.021*
Italian	-0.043***	0.034**
Education, ref: Compulsory		
N.A.	0.076***	0.036**
Secondary	-0.096***	-0.054***
College	-0.357***	-0.412***
Plant area, ref: North-west		
North-east	0.099***	0.062***
Center	0.069***	0.071***
South	0.345***	0.259***
Islands	0.308***	0.243***
Temporary Contrat, ref: Fixed-term		
Apprenticeship		-0.962***
Freelance		-0.180***
Internship		-0.048*
Other		0.225
Working time, ref: Full-time		
N.A.		-0.341***
Part-time		-0.087***
Controls for:		
Occupation (9)	No	Yes
Sector (17)	No	Yes
N.	60153	
Failures	40888	
LR χ^2	2123.89	9619.40

13 The last two quarters of our period show lower hazard of termination even because of the right censoring.

4 Conclusive Remarks

The analyses focused on the job market entry of young individuals in Italy with respect to the temporary or permanent nature of their jobs. The tightness in the Italian labour market for new entrants and young workers as well as their high propensity to get temporary jobs suggest to deepen this subject.

Using the new CC data on contracts both in the public and the private sector, we exploited the detailed flows of information on jobs evolution over the period 2008–2010, accounting for the transformation of temporary jobs into permanent ones, the extension and the early termination of temporary jobs.

Young individuals have been selected from our CC sample dataset to study their probability to get an open-ended contract and to move from a temporary job to a permanent one as well as their temporary jobs durations.

In Italy the probability to get a permanent job is low and has been affected by the crisis. We found that it increases with age, as the probability to move from a temporary job to a permanent one. Therefore, the chances to get an open-ended contract seem to have been mostly offered to those individuals with more experience, which also have the long lasting temporary jobs.

On the other hand, young workers who have spent many years on education have low chances to start with a permanent job, however their temporary jobs are more easily transformed into permanent ones. More educated individuals usually working on temporary basis are engaged in long-lasting job relationships.

With respect to other types of temporary jobs, fixed-term contracts show the highest chance of transformation in permanent contracts. However, the estimates of the survival probability suggest that their expected length is the lowest. The apprenticeship contracts show the highest effective job duration.

References

Berton F., M. Richiardi e S. Sacchi. 2009a. Flex-Insecurity - Perchè in Italia la flessibilità diventa precarietà. Bologna.: Il Mulino

Berton F., M. Richiardi e S. Sacchi. 2009b. Are Temporary Jobs a Port of Entry into Permanent Employment? Evidence from Matched Employer-Employee Data, Working Papers 6, University of Torino, Department of Economics and Public Finance "G. Prato"

Booth A.L., M. Francesconi e J. Frank. 2002. Temporary jobs: stepping stones or dead ends? The Economic Journal, vol. 112, pp. 189 - 213

Coccia G., Rossi B. (2010) I giovani in Italia, Ministero del Lavoro e delle Politiche Sociali.

Gagliarducci S. 2005. The dynamics of repeated temporary jobs, Labour Economics, 12, 429-448

Gomes P. 2012. Labour market flows: facts from the United Kingdom, Labour Economics, 19, 165-175

Picchio M. 2008. Temporary Contracts and Transitions to Stable Jobs in Italy, LABOUR 22(s1), pp. 147-174

Staffolani S. e M. Lilla. 2009. La disuguaglianza e la volatilità nei redditi di lavoro: giovani e anziani, operai e impiegati; in L'Italia delle disuguaglianze a cura di L. Cappellari, P. Naticchioni e S. Staffolani, Carocci, Roma

Strano G., T. Lang, B. Rossi e V. Sorci. 2010. Il sistema delle comunicazioni obbligatorie: uno strumento per l'analisi del mercato del lavoro, Ministero del Lavoro e delle Politiche Sociali

Norme redazionali

La Rivista di Statistica Ufficiale pubblica contributi originali nella sezione “Temi trattati” ed eventuali discussioni a largo spettro nella sezione “Interventi”. Possono essere pubblicati articoli oggetto di comunicazioni a convegni, riportandone il riferimento specifico. Gli articoli devono essere fatti pervenire al Comitato di redazione delle pubblicazioni scientifiche Istat corredati, a parte, da una nota informativa dell’Autore contenente: appartenenza ad istituzioni, attività prevalente, qualifica, indirizzo, casella di posta elettronica, recapito telefonico e l’autorizzazione alla pubblicazione firmata dagli Autori. Ogni articolo prima della pubblicazione dovrà ricevere il parere favorevole di un referente scelto tra gli esperti dei diversi temi affrontati. Gli originali, anche se non pubblicati, non si restituiscono.

Per l’impaginazione dei lavori gli autori sono tenuti a conformarsi rigorosamente agli standard editoriali fissati dal Comitato di redazione e contenuti nel file Template.doc disponibile on line o su richiesta. In base a tali standard la lunghezza dei contributi originali per entrambe le sezioni dovrà essere limitata entro le 30-35 pagine.

Tutti i lavori devono essere corredati di un sommario nella lingua in cui sono redatti (non più di 12 righe); quelli in italiano dovranno prevedere anche un Abstract in inglese. La bibliografia, in ordine alfabetico per autore, deve essere riportata in elenco a parte alla fine dell’articolo. Quando nel testo si fa riferimento ad una pubblicazione citata nell’elenco, si metta in parentesi tonda il nome dell’autore, l’anno di pubblicazione ed eventualmente la pagina citata. Ad esempio (Bianchi, 1987, Rossi, 1988, p. 55). Quando l’autore compare più volte nello stesso anno l’ordine verrà dato dall’aggiunta di una lettera minuscola accanto all’anno di pubblicazione. Ad esempio (Bianchi, 1987a, 1987b).

Nella bibliografia le citazioni di libri e articoli vanno indicate nel seguente modo. Per i libri: cognome dell’autore seguito dall’iniziale in maiuscolo del nome, il titolo in corsivo dell’opera, l’editore, il luogo di edizione e l’anno di pubblicazione. Per gli articoli: dopo l’indicazione dell’autore si riporta il titolo tra virgolette, il titolo completo in corsivo della rivista, il numero del fascicolo e l’anno di pubblicazione. Nei riferimenti bibliografici non si devono usare abbreviazioni.

Nel testo dovrà essere di norma utilizzato il corsivo per le parole in lingua straniera e il corsivo o grassetto per quei termini o locuzioni che si vogliano porre in particolare evidenza (non vanno adoperati, per tali scopi, il maiuscolo, la sottolineatura o altro).

Gli articoli pubblicati impegnano esclusivamente gli Autori, le opinioni espresse non implicano alcuna responsabilità da parte dell’Istat.

La proprietà letteraria degli articoli pubblicati spetta alla Rivista di statistica ufficiale.

E’ vietata a norma di legge la riproduzione anche parziale senza autorizzazione e senza citarne la fonte.

Per contattare il Comitato di redazione delle pubblicazioni scientifiche Istat e per inviare lavori: rivista@istat.it. Oppure scrivere a:

Segreteria del Comitato di redazione della Rivista di Statistica Ufficiale
All’attenzione di Gilda Sonetti
Via Cesare Balbo, 16
00184 Roma

La Rivista di Statistica Ufficiale accoglie lavori che hanno come oggetto la misurazione e la comprensione dei fenomeni sociali, demografici, economici ed ambientali, la costruzione di sistemi informativi e di indicatori come supporto per le decisioni pubbliche e private, nonché le questioni di natura metodologica, tecnologica e istituzionale connesse ai processi di produzione delle informazioni statistiche e rilevanti ai fini del perseguimento dei fini della statistica ufficiale.

La Rivista di Statistica Ufficiale si propone di promuovere la collaborazione tra il mondo della ricerca scientifica, gli utilizzatori dell'informazione statistica e la statistica ufficiale, al fine di migliorare la qualità e l'analisi dei dati.

La pubblicazione nasce nel 1992 come collana di monografie "Quaderni di Ricerca ISTAT". Nel 1999 la collana viene affidata ad un editore esterno e diviene quadrimestrale con la denominazione "Quaderni di Ricerca - Rivista di Statistica Ufficiale". L'attuale denominazione, "Rivista di Statistica Ufficiale", viene assunta a partire dal n. 1/2006 e l'Istat torna ad essere editore in proprio della pubblicazione.

1B012012001000000

ISSN 1828-1982

€ 10,00